# Early word order and animacy

Cannizzaro, Courtney Leigh

Early word order and animacy

Groningen Dissertations in Linguistics 104
ISSN: 0928-0030
ISBN: 978-90-367-5640-2
e-ISBN: 978-90-367-569-38

This investigation was supported by a grant from the Netherlands Organisation for Scientific Research, NWO, awarded to Petra Hendriks (grant no. 277-70-005) for the VICI project "Asymmetries in Grammar"

The work presented here was carried out under the auspices of the Netherlands Graduate School of Linguistics (Landelijke Onderzoeksschool Taalwetenschap) and the Center for Language and Cognition Groningen of the Faculty of Arts of the University of Groningen. Research reported on English-speaking participants was made possible in part by the Institute for Research in Cognitive Science at the University of Pennsylvania and the Stichting Groninger Universiteitsfonds.

Printed by Wöhrmann Print Service, Zutphen
Cover art by Esther Ris

RIJKSUNIVERSITEIT GRONINGEN

# Early word order and animacy

Proefschrift

ter verkrijging van het doctoraat in de
Letteren
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
donderdag 13 september 2012
om 12:45 uur

door

**Courtney Leigh Cannizzaro**

geboren op 6 juli 1982
te New Orleans, USA

| | |
|---|---|
| Promotor: | Prof.dr. P. Hendriks |
| | |
| Beoordelingscommissie: | Prof.dr. J. Hoeksema<br>Prof.dr. H. de Hoop<br>Prof.dr. J. Trueswell |

*To Mom, Dad, and Gavin*

# Acknowledgements

As a member of the "Asymmetries in Grammar" NWO Vici Project, I have been privileged to work with a fantastic team of researchers: Petra Hendriks, Charlotte Koster, Bart Hollebrandse, John Hoeks, Catharina Hartman, Sanne Kuijper, Jacolien van Rij and Peter de Swart. The project would not have been possible without grant recipient and project leader Petra, who kept the group energized with her enthusiasm. I am incredibly grateful for her firm but friendly approach to supervising, which kept me on track and thinking positively. I am also thankful to the rest of the team for their hours of collaborative brainstorming and insightful feedback. I am particularly indebted to Jacolien for the time she has spent giving me a crash course in R scripting and for her readiness to discuss my questions on a daily basis. Her natural drive to carry out quality research never ceased to be inspiring. Peter should also be sincerely thanked for the time he volunteered as my "animacy mentor" during his time in Groningen, which involved, among other things, reading much of my manuscript. The feedback I received from the regular attendees of the Groningen Acquisition Lab meetings was also much appreciated; among them are Angeliek van Hout, Jennifer Spenader, Emar Maier, Ruggero Montalto, Evelien Krikhaar, and Susanne Grassmann. My reading committee must also be thanked for the time they took to read and evaluate this work: Jack Hoeksema, Helen de Hoop, and John Trueswell.

Anyone who works in the west corridor of the fourth floor knows that room 418 served as the Asymmetries project's central station. This is where you could find three of the project members cozily occupying a two-person room, a highly frequented filing cabinet with a treasure trove of essential keys, a swarm of student assistants and interns, a massive rescued cactus, and most importantly, a shelf full of cookies and *drop* candy. In addition to nurturing the healthy sense of community we had as a greater research group, these tiny quarters were conducive to the formation of deep friendship between me and my officemates Charlotte and Sanne. I value the afternoon conversations I had with Charlotte—also from the United States South—who came to be much more than only a supervisor to me. Sanne's unwavering support in all aspects of my life has been invaluable. If it were not for Charlotte and Sanne (and their loving husbands), I do not know who would have lent me drills, helped me move, or served as my emergency contacts. Simply put, they are family to me.

Important contributors not only to the *gezellige* atmosphere of the Asymmetries Project but also to its powerhouse of productivity have been the hard-working student assistants, interns, and volunteers. I want to thank Saskia van den Akker, Sanne Berends, Frederike Groothoff, Sabine van der Ham, Laura Hemstra, Merel Heerink, Trudy Krajenbrink, Kaitlin Mignella, Jessica Overweg, Alma Veenstra, Maaike Veeninga, Ellis Wieringa, Laura Wismans, Edgar Weiffenbach, and Fransje van Weerden for their help with my project. It was truly a pleasure to work with them, especially when testing the adorable, but unpredictable preschoolers. I do not take for granted how much I was able to rely on these individuals for high-quality and enthusiastic assistance.

I want to thank both the adult and child participants who came to the lab to be tested. I appreciate the time that interested parents took to bring their children to the lab. It was very enjoyable for me to chat with them about their children and my research. I also want to thank those children who participated in the act-out task at day cares in the Groningen area and their parents who gave permission. Rather than getting to meet the parents, I was instead able to chat with the staff of the day cares during morning fruit and song time. Thank you to the cooperating day cares: BSO Schildersbuurt, Kinderdagverblijf SKSG de Regenboog, Kinderdagverblijf de Wikke, Kinderdagverblijf SKSG de Rode Pimpernel, Kinderdagverblijf de Boomhut, Lutje Potje Peuterspeelzaal, and Kinderopvang Hamertje-Tik.

Concerning the practical matter of testing children at the university, it helped that we had cooperative porters in the Harmony Building and a hospitable Frisian department (where the Eye Lab is located). Essential to experimentation was properly operating equipment, so many thanks to the Audio Visual Services (especially Jan-Willem Pomper and Callista Jippes) and to the CIT's "Vincent, Vincent & Bert" (Vincent Hoekzema, Vincent Manisuwa, and Bert Gernaat). Those who created the custom animations displayed on the eye tracker should also be thanked: Rik Schlimbach and Liske van der Vliet. Though Rik and Liske were art students with no experience with creating material for psycholinguistic experiments, they amazed me with their patience and eventual understanding of exactly what I needed. The animations were based on illustrations created for a pilot carried out by a group of bachelor students at the Hanzehogeschool in Groningen: Valerie Dubois, Machteld Kort, Cynthia Werkman, students of Margreet Luinge.

I would not have acquired the skills critical for designing a successful eye tracking study had it not been for help from members of the European eye tracking community, particularly Frauke Berger, Christina Bergmann, Oda-Christina Brandt, Tom Fritszsche, Silvana Poltrack, and Barbara Höhle. Thanks to them, tricky things quickly became *clear* (winky face). I also received invaluable advice about eye tracking from the members of the Institute of Research in Cognitive Science (IRCS) at the University of Pennsylvania, where I had a six-week visit in the spring of 2010 and a shorter follow-up visit in the spring of 2011. Most notably, I got feedback from John Trueswell, who helped me to greatly improve my experimental design and technique for data analysis. It was very kind of him to let me use one of the eye trackers in his lab to test children at day cares in Philadelphia. Together with John's extremely friendly and organized research assistant Alon Hafri, I accomplished a lot in a short amount of time. I also want to sincerely thank the other members of the IRCS for their assistance and hospitality, especially Laurel Sweeney, Lila Gleitman, Adrienne (and Shaune and Penny) Scutellaro, Joe Fruhwald, Kyle Gorman, Tamara Nicol Medina, Ben Armstrong, and Ann Bunger.

I have also received general support from further members of the linguist and psychologist communities. Regular meetings with our group's OT "cousins" in Nijmegen were always fruitful, usually involving discussions with Helen de Hoop, Lotte Hogeweg, Geertje van Bergen, Kees de Schepper, and Sander Lestrade. I must also thank Monique Lamers, Henriëtte de Swart, Paula Fikkert, Sharon Unsworth, Esther Ruigendijk, Eve Clark, Jill DeVilliers, Géraldine Legendre, Irina Sekerina, Jennifer Arnold, and Florian Jaeger for their helpful comments at conferences and workshops over the years.

I would like to thank all of the members of the CLCG, coordinator Wyke van de Meer, and the secretaries of the Dutch Department. The CLCG is a pleasant place to do research, with an extremely approachable senior research staff. The following CLCG AIOs should be specifically mentioned for their assistance—from showing me how to create a user script in E-Prime to helping with babysitting the younger siblings of my experimental participants: Aysa Arylova, Veerle Baaijen, Çağrı Çöltekin, Myrthe Gosen, Marlies Kluck, Hanneke Loerts, Zhenya Markovskaya, Barbara Plank, Martijn Wieling, and Ankelien Schippers. The availability of these and other AIOs after hours should not go un-thanked, for *borreling* is an essential activity that cannot be done alone. Regulars were Daniel de Kok, Dörte Hessler, Gideon Kotze, Karin Beijering, Tim Kallenborn, and Ryan Taylor. Erik Tjong Kim Sang, my fellow pinball enthusiast, deserves special thanks for attending each of my birthday gatherings with a sack full of coins. In addition to the staple borreling, the following individuals regularly joined me in indulgent shopping, traveling, and gossiping, (respectively): Ildikó Berzlánovich, Tim van de Cruys, and Radek Šimík. Finally, one gentleman who bore the burden of being practically my only non-linguist friend in Groningen deserves recognition. If anyone can help me snack my worries away, it is my clever old Oosterpoort neighbor Wessel Roose.

My general success with language and linguistics may not have been possible were it not for my grade school English teacher, Jane Butera McGovern (the very woman who taught me how to use *were* to convey the subjunctive!). Perhaps it did not have to be her (or rather *she*) who taught me what a noun and a verb are for me to have fallen in love with linguistics already in sixth grade, but it was, so I thank her sincerely.

Which leaves my family. My brother has played a fundamental role in my life. I have learned from him how to be a critical thinker, which has helped me succeed both as an academic and as a citizen of the world. He and his beautiful family are a consistent source of warmth to me in spite of the distance between us. Finally, the unconditional support I have always received from my parents lies at the root of all of my accomplishments. I would like to thank them for the steady flow of encouragement over the last years. I needed it!

x

# Contents

# Abbreviations

| | |
|---|---|
| NP | noun phrase |
| S-O | subject-object (word order) |
| C&M | Chapman & Miller (1975) |
| OT | Optimality Theory |
| M >> F | markedness outranks faithfulness (OT constraints) |

## Types of languages

| | |
|---|---|
| SVO, SOV, VSO, VOS, OVS, OSV | subject/verb/object constituent order |
| V2 | verb-second |

## Types of sentences

| | |
|---|---|
| S=O | subject and object equal on animacy hierarchy |
| S>O | subject higher than object on animacy hierarchy |
| S<O | subject lower than object on animacy hierarchy |
| S≠O | subject and object not equal in animacy (S<O or S>O) |
| | |
| [+an +an] | animate-animate |
| [+an -an] | animate-inanimate |
| [-an +an] | inanimate-animate |
| [-an -an] | inanimate-inanimate |

## Word order in comprehension and production

| | |
|---|---|
| SO interpretation | sentence interpreted as SO |
| OS interpretation | sentence interpreted as OS |
| | |
| SO order | sentence produced as SO |
| OS order | sentence produced as OS |

## Child language terms

| | |
|---|---|
| 1;8 | 1 year and 8 months |
| mo. | month |
| MLUm | mean length of utterance (in morphemes) |
| | |
| CDI | MacArthur-Bates Communicative Development Inventory |
| N-CDI | Dutch (*Nederlandse*) Communicative Development Inventory |
| KINT | *Kleuter Inventarisatie Nederlandse Taalverwerving* |

## Experiment terms

| | |
|---|---|
| ERP | event-related potentials |
| RT | reaction time |
| VOL | voice onset latency |
| AOI | areas of interest |

## Glosses

| | |
|---|---|
| 1,3 | first, third person |
| ACC | accusative |
| FAC | factive |
| FIN | finite |
| INF | infinite |
| M | masculine |
| F | feminine |
| NOM | nominative |
| OBJ | object |
| PL | plural |
| PST | past |
| PTC | participle |
| SNG | singular |
| SU | subject |

# 1 Introduction

## When cars push cows

If you are a native speaker of English, sentence (1) probably invokes for you an image from a horror movie about a fast food joint.

(1) The hamburger is eating the boy

Word order in English is so strict, that the first noun phrase *the hamburger* must be the subject, or the one who is eating, and the second noun phrase *the boy* must be the object, or the one who is being eaten. As a result, you are forced to imagine a big hamburger with teeth and a gaping mouth in which a small boy could fit. Even though I am acquainted with German and Dutch, in which word order is more flexible, I hear this sentence as a native English-speaker and immediately fear for the boy.

It is possible that as children, we were less sensitive to word order and would have allowed a different interpretation of a sentence like (1). Despite the vivid imaginations of children, very young English-speakers could believe (boringly) that the boy is the one who is doing the eating. This is not entirely strange given that the event represented by (1) when word order is ignored is more likely to occur than if word order were strictly honored. After all, *eat* is not neutral with regard to the arguments it can take. It requires a living or *animate* subject, thereby making the boy the best "eater" of the two noun phrases given.

Now consider a sentence like (2), which is fully reversible in that both cars and cows are capable of pushing, and no stereotypical event is represented by either possible interpretation.

(2) The car is pushing the cow

It seems that when presented with a sentence like (2) and asked to act it out with toys, a child between the age of two-and-a-half and three-and-a-half will be about 50% likely to make the cow push the car (Chan, Lieven, & Tomasello, 2009; Chapman & Miller, 1975). Even *without* the influence of verb argument requirements or event likelihood, English-speaking children appear to regularly ignore word order in favor of interpretations that allow animate entities to act upon inanimate entities.

This phenomenon in which preschoolers fail to reliably use word order when interpreting reversible transitive sentences is particularly intriguing when set against the backdrop of their apparent adult-like production of word order (Angiolillo & Goldin-Meadow, 1982; Chapman & Miller, 1975; McClellan, Yewchuk, & Holdgrafer, 1986). This is striking since it is more common that comprehension exceeds production. Think about how children who cannot yet speak seem to understand much of what adults say to them. Think about how you can understand your foreign language teacher, but the quality of your own speech in the language pales in comparison. While any asymmetry between production and comprehension deserves attention, the instance of production exceeding comprehension is certainly more extraordinary.

The remarkable possibility that children employ word order in a reversible sentence like (2) differently in comprehension than they do in production inspired the two central themes of this dissertation: asymmetry and animacy. Do children produce subject-object word order more successfully than they are able to understand it? And if so, is their comprehension systematically affected by the relative animacy of subject and object?

In order to answer these questions, I establish which theoretical framework I believe is best able to account for asymmetric performance by young children. I also discuss how the comprehension and production of word order can most effectively be tested in preschoolers. Using a variety of comprehension tasks, I investigate how Dutch- and English-speaking preschoolers, as well as adult controls, interpret reversible transitive sentences in which the relative animacy of subject and object have been manipulated. For each comprehension task, an accompanying production task determines how the same transitive sentences are formed by the same individuals. The pattern of answers given on these tasks are analyzed, in addition to gaze data collected via automated eye tracking during sentence comprehension and production. Based on the results of these experiments, I conclude whether there is an asymmetry between production and comprehension and whether comprehension is affected by animacy.

In the rest of this introduction, I identify the implications this research has for acquisitional and grammatical theory. I then outline some important terminology and assumptions relevant to the current discussion. I end with an overview of the chapters that follow.

## 1.1 Acquisition of word order

As opposed to, say, the acquisition of partitive case-marking, evident in only a handful of languages, the acquisition of agent-patient relations is universally relevant. *Agent-patient relations* simply refers to the "who does what to whom", with the agent as the "who" and the patient the "whom." All children must eventually come to encode and decode agents and patients in their target language, so it can be assumed that just when and how they do this will reveal something about the universal function and acquisition of grammar.

If young children encode agent-patient relations better than they are able to decode the same relations, such an asymmetry must be accounted for in acquisitional and grammatical theory. In this section, I first provide a rough sketch of the main themes in acquisition research before making the assertion that a grammatical explanation is preferred over an extra-grammatical explanation of a word order asymmetry. I then briefly discuss in what way a grammatical explanation of this asymmetry has implications for acquisitional theory and for grammar in general.

### 1.1.1 Competence, performance, and continuity

Language acquisition research carries a traditional distinction between *competence* and *performance*, first made by Chomsky (1965). The distinction has been characterized in somewhat different terms, depending on whether one views competence or performance as central in acquisition research. Charles Yang, who stresses the importance of competence, defines performance as what keeps us from witnessing underlying competence: "Language use is the composite of linguistic, cognitive and perceptual factors many of which, in the child's case, are still in development and maturation. It is therefore difficult to draw inferences about the learner's linguistic knowledge from his linguistic behavior" (Yang, 2010: 1). Under this point of view, linguistic performance is believed to obscure what researchers should be interested in determining, i.e. children's underlying linguistic competence.

In contrast, Elizabeth Bates and Brian MacWhinney, who stress the importance of performance, define competence as "the abstract knowledge of language possessed by an ideal speaker-listener, removed from the constraints and inconveniences of real-time language use" (Bates & MacWhinney, 1989: 3). That their use of "inconveniences" is meant sardonically becomes clear when they go on to define performance as "the actual process of language use by real people in real situations." Under this view, linguistic competence is less key than the practical application of knowledge put to use in everyday communication, i.e. performance.

The distinction between a competence approach and performance approach goes hand-in-hand with the traditional dichotomy in language acquisition research: nativists/generativists versus functionalists/constructivists. Nativists emphasize how quickly and seemingly effortlessly children acquire language at all odds, and without direct instruction. They often advocate a separate, biologically endowed module in the brain specific to language. Nativists have proposed a quick system to language learning involving the simple setting of a (usually) binary parameter upon hearing a modest amount of input (Chomsky & Lasnik, 1993). Functionalists, on the other hand, emphasize the effort required of children to learn their language, as well as the importance of child-directed speech and indirect correction (Bates & MacWhinney, 1989; Tomasello, 1992). They believe that language results from the interaction of general cognitive abilities and information from the environment. These functionalist explanations attribute children's mistakes to a slow language learning process that requires many examples in the input before linguistic rules can be generalized beyond specific lexical items.

A key task of the acquisitionist, whether nativist or functionalist, is to address why child language differs from adult language. Supporters of the *continuity hypothesis* claim that the underlying system of child grammar is the same as that of the adult (Crain, Goro, & Minai, 2007; Thornton, 2002). Some believe that this continuity further includes processing mechanisms (Clahsen & Felser, 2006). A widely accepted version of the continuity hypothesis holds that a child's language will differ from the target language only in ways that adult languages are able to differ from each other. This belief follows from a theory of universal grammar, originating with generativists, which assumes that linguistic knowledge consists of universal rules or constraints that underlie all languages (Chomsky & Lasnik, 1993; Prince & Smolensky, 1993). A child's language at any given point would, then, at least be abiding by some rule that is allowed in one of the world's languages. Opponents of the continuity hypothesis, often functionalists, believe that the language of children is much more concrete and item-based than the language of adults, even if the two systems result in similar behavior (Tomasello, 2000).

While nativists and functionalists disagree about continuity between the linguistic systems of adults and children, both have been known to call upon extra-grammatical explanations of non-adult-like linguistic behavior of children. *Extra-grammatical* refers to mechanisms believed to be outside the grammar, like pragmatic or cognitive abilities. Nativists may call upon such explanations in the interest of maintaining continuity, and the functionalists may do so to support a separation of language from general cognitive biases.

### 1.1.2 Grammatical explanation

As a starting point, I assert that word order asymmetry should be investigated in terms of a theoretical framework that allows for a grammatical explanation. First of all, word order is generally accepted to be a product of the grammatical, syntactic rules of a language. Furthermore, extra-grammatical explanations pose the risk of being too easily called upon whenever needed. Finally, there is a grammatical framework available, namely Optimality Theory, that takes into account differences between production and comprehension, as well as effects of animacy on sentence interpretation.

Optimality Theory can be seen as a middle ground between nativism and functionalism. Like nativism, it attributes innate knowledge in the form of universal grammar to the child, but does so in the form of soft, not hard and inflexible rules. Like functionalism, it acknowledges the importance of input in the gradual learning process, but at the same time supports the idea of continuity—that child language can be described in terms of adult language. This framework is nicely able to maintain continuity by (i) distinguishing between the task of a grammar during comprehension versus during production, and (ii) incorporating what is considered to be a general cognitive bias (e.g. animacy) into the grammar itself. The implications of each of these capabilities is discussed in turn below. Chapter 2 offers a more detailed description of the Optimality Theoretic framework as well as an outline of alternate explanations.

#### *1.1.2.1 Asymmetric grammar*

Because it is so common for children to understand more than they are able to produce themselves, the traditional view of the relationship between comprehension and production is that receptive language exceeds and precedes language production (Bates, Dale, & Thal, 1995; Clark & Hecht, 1983; Fraser, Bellugi, & Brown, 1963; Ingram, 1974). But we know that the asymmetry can also occur in the opposite direction. A well-known example of production exceeding comprehension is children's use of pronouns. Children until the age of about six incorrectly allow a reflexive (*himself*) interpretation of the pronoun *him* in languages such as English, Dutch, Italian, and French (Chien & Wexler, 1990; Hamann, Kowalski, & Philip, 1997; Koster, 1993; McKee, 1992; Philip & Coopmans, 1996; Sekerina, Stromswold, & Hestvik, 2004). However, children correctly produce both reflexives and pronouns in their own speech at an early age (Bloom, Barss, Nicol, & Conway, 1994; Spenader, Smits, & Hendriks, 2009; J. G. de Villiers, Cahillane, & Altreuter, 2006). There have been similar findings for constrastive stress (Kuijper & Groothoff, 2010) and indefinite objects (Unsworth, 2007).

How genuine—that is, how much due to competence, or the grammar—an asymmetry is viewed within a theory is determined by which aspects of comprehension and production are delegated to the grammar and which are delegated to general cognition (Hendriks & Koster, 2010). Despite observed asymmetries in child language, children's grammatical knowledge is often considered to be symmetric between comprehension and production. The source of developmental asymmetry under such a view must lie outside the grammar. Regarding non-adult-like production, for example, non-target pronunciations may be attributed to immature motor skills; likewise, subject omissions (*want cookie*) may be attributed to limited processing capacity (Valian, 1991) or pragmatic limitations (Weissenborn, 1992). Regarding non-adult-like interpretation, the problem with pronouns like *him* has been attributed to processing limitations (Reinhart, 2006) or experimental artifacts (Conroy, Takahashi, Lidz, & Phillips, 2009). As an alternative approach, Optimality Theory accounts for both types of asymmetry by incorporating the different demands of comprehension and production into the grammar itself. This framework eliminates the need to call upon extra-grammatical explanations for non-adult-like linguistic behavior of children.

#### *1.1.2.2 Animacy in the grammar*

The role that animacy may play in the word order asymmetry can be seen as difficult to define for two reasons. First of all, as will be shown in Chapter 2, little research has aimed to isolate the effect of animacy on word order in comprehension or production. Research on English and other languages has often explored the acquisition of subject-object word order in light of combinations of the following types of information: case or agreement inflection, word order, animacy, and event probability (Chapman & Kohn, 1978; Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008; Hakuta, 1982; Koff, Kramer, & Fowles, 1980; Lindner, 2003; Pléh, 1989; Schaner-Wolles, 1989; Sinclair & Bronckart, 1972; Slobin & Bever, 1982; Sokolov, 1989; Strohner & Nelson, 1974; Turner & Rommetveit, 1967). Because each of these factors indeed plays a role in the interpretation of transitive sentences, it is crucial that all information other than animacy and word order be kept constant in order to determine whether animacy alone will cause reversed interpretations of transitive sentences.

The second difficulty arises from the apparent influence of animacy in children's comprehension, making a fully grammatical account of S-O word order problematic for traditional theories. The few studies that have isolated the effects of animacy and word order on comprehension (Chan et al., 2009; Chapman & Miller, 1975) suggest that preschoolers do give reversed interpretations of transitive sentences with an inanimate subject and animate object. In order to make reference to this influence of animacy, a theory would have to define the preference for animate subjects and inanimate objects as *linguistic* in nature—a preference that is traditionally considered to be a general, *non-linguistic* bias of cognition. This issue is touched on by MacWhinney and colleagues (MacWhinney, Bates, & Kliegl, 1984: 137):

> The psycholinguistic literature is rich in studies demonstrating a probabilistic relation between animacy and agency in English. This relation involves a tendency to perceive the more animate of two nouns as the agent of an action [ . . . ] The existence of this tendency is widely recognized, but it remains to be seen whether semantic cues of this type should be regarded as a systematic part of the grammar [ . . . ] The fact that animacy plays such an obvious role in the grammar of some languages should at least leave us open to the possibility that a similar semantic distinction could operate in a less absolute way in other languages.

While these authors acknowledge the cross-linguistic effect of animacy, they continue to doubt how much weight animacy should be given in universal grammar. Lamers and de Hoop (2004, 2005), on the other hand, view the cross-linguistic evidence that animacy has been grammaticalized in some languages as motivation to give animacy an official role in their Optimality Theoretic model of grammar. The incorporation of animacy into the grammar allows for a continuous, grammatical account of the acquisition of S-O word order (Hendriks & Koster, 2010; Hendriks, de Hoop, & Lamers, 2005), described in detail in the next chapter.

## 1.2 Terminology and assumptions

Simplified terminology and crucial assumptions of the current discussion are specified here. I first clarify what will be referred to as *word order* and what will be assumed about how word order in comprehension and production are linked. I then define an animacy hierarchy that is adequate for the current discussion, identifying what will be referred to as *animate* and *inanimate*. Finally, I make the crucial assumption that psycholinguistic experiments can tell us something not only about linguistic knowledge, but also about how comprehension compares to production.

### 1.2.1 Simplified word order

There are numerous types of word order that a child must acquire. The order of constituents (e.g. verbs, verb-arguments, or adverbs) in a particular language usually vary between declaratives, questions, subordinate clauses, relative clauses, and clauses with negation or in the passive voice. The present study investigates only the order of subject and object in simple, active, declarative sentences used with transitive verbs, like sentence (2), which we have seen already.

(2) The car is pushing the cow

The term *word order* used here refers to this subject-object (S-O) word order, unless otherwise specified. *Early word order* refers simply to word order in children under the age of four, or *preschoolers*.

Two levels of linguistic representation are relevant for this study: a surface level (the form) and an underlying level (the meaning). Elements at the surface level of sentence (2) are shown twice in Figure 1.1. In both comprehension and production of English and Dutch it is assumed that in simple, active, declarative sentences, the first NP is the subject *S* and the second NP is the object *O*. Elements of the underlying level reflect a meaning, expressed in Figure 1.1 as PUSH <x, y>, where *x* is the agent and *y* is the patient. It is assumed that for English and Dutch that in simple, active, declarative sentences, the *x* corresponds directly to the surface subject and the *y* corresponds directly to the surface object, whether in comprehension or production.

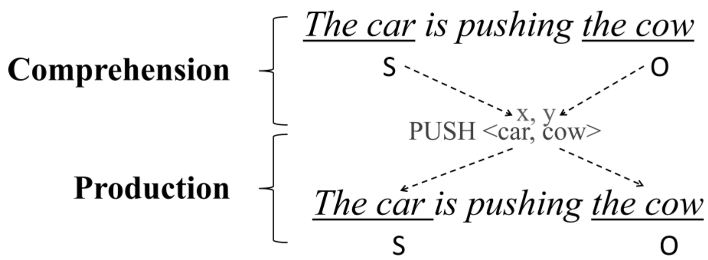


**Figure 1.1 Underlying and surface levels of language comprehension and production**

This one-to-one correspondence between the surface and underlying elements is adequate in the current investigation. It would not suffice in a study that included, for example, passive sentences, in which the patient becomes the surface subject and the agent becomes an oblique (*The cow is pushed by the car*). For active sentences, however, this direct mapping can be used simply to determine if children are mapping *S* to *x* and *O* to *y* in comprehension and *x* to *S* and *y* to *O* in production. That is, if a child expresses the meaning PUSH <car, cow> with the sentence "The car is pushing the cow", then the word order was used in an adult-like way. But if the child hears the very same sentence and interprets it as meaning PUSH <cow, car> then the order was not used in an adult-like way. In comprehension, order is described as either a *SO interpretation* or *OS interpretation*; in production, order is described as either having *SO order* or *OS order*.

### 1.2.2 Simplified animacy hierarchy

Animacy is considered one of several "prominence" factors, such as familiarity, topic, definiteness, and person, that help determine morphological and syntactic structure in a language (Allan, 2008). The entire set of factors can either be incorporated into single, comprehensive hierarchies, such as that of Silverstein (1976) or Wilkins (Van Valin & Wilkins, 1996; Wilkins, 1990), or they can be reduced to three distinct scales of grammatical prominence: definiteness, person, and animacy.

The most simplified individual animacy hierarchy is human >> non-human, but it is more common to involve at least a three-way distinction: human >> animal/animate >> inanimate (Aissen, 2003; Comrie, 1989). Fine-grained hierarchies may include many distinctions (humans, organizations, animals, intelligent machines, vehicles, concrete/non-concrete inanimates, place, and time), which is useful when coding corpora for automatic parsing (Zaenen et al., 1976). Wilkins' comprehensive prominence hierarchy actually reserves the terms *animate* and *inanimate* for living entities, while non-living entities are referred to as either *motive* or *non-motive* (Van Valin & Wilkins, 1996; Wilkins, 1990).

In order to find a common ground between the simplified and detailed animacy hierarchies, I maintain only four distinctions necessary for the current discussion. I therefore use the following hierarchy, similar to that used by Corrigan (1986):

ANIMATE (humans >> animals) >> INANIMATE (vehicles >> stationary entities)

Accordingly, humans and animals will be referred to as *animate* and vehicles and stationary entities will be referred to as *inanimate*. Furthermore, it is assumed that humans are more animate than animals, and stationary entities are more inanimate than vehicles.

### 1.2.3 Psycholinguistic experimentation

An important assumption made in any psycholinguistic investigation is that an experiment can test linguistic knowledge. In their book describing methods for assessing the syntactic knowledge of children, McDaniel, McKee, & Cairns (1996: xiv-xv) propose that this is a viable assumption and that it can be applied in acquisition research. Statistical methods and experimental tools have been established over decades to help researchers more confidently distinguish psychologically real phenomena from "noise." Statistics are used to determine where there is systematic variation, and a carefully controlled experimental design helps eliminate with some certainty which factors are not the source of the variation. In this way, psycholinguistic experimentation is believed to be able to reveal something about linguistic behavior and, ultimately, about underlying linguistic knowledge.

Thus, I assume that the tasks and design I use, described in detail in Chapter 3 and applied in Chapters 4 – 7, are quantitative tools sophisticated enough to test the linguistic knowledge of adults and children. I additionally assume that the degree to which word order is successfully used in comprehension and production measured in an experiment can fairly be compared, which I do in Chapter 8. By testing the same group of participants and using the same set of materials in both tasks, the fairness of this comparison is optimized.

## 1.3 Overview of chapters

This dissertation can be broken down into three sections: (i) background, (ii) experiments, and (iii) discussion and conclusions. In the two background chapters that follow, the phenomenon of word order asymmetry is introduced, and the model and methodologies are described that are used to investigate it. In Chapter 2, I review previous studies in which word order has been tested in preschoolers, revealing the intriguing possibility of a word order asymmetry. I then evaluate acquisitional theories and ultimately deem the constraint-based model Optimality Theory as best able to model asymmetry. I introduce Hendriks, de Hoop, and Lamers' (2005) model of word order asymmetry and de Hoop and Lamers' (2006) model of real-time grammar application. Making use of these models, I predict that Dutch- and English-speaking preschoolers will be able to use word order correctly, but will fail to interpret word order in the face of conflicting animacy information in comprehension. In Chapter 3, I describe methods of testing production and comprehension in preschoolers, with a focus on elicitation, act-out, picture selection, preferential looking tasks, as well as the incorporation of eye tracking measurements. I describe materials in detail and emphasize the importance of a controlled design.

In the five experiment chapters, I report six experiments that test the comprehension and production of word order. An overview of the experiments appears in Table 1.1. Note that a-experiments test comprehension and b-experiments test production.

**Table 1.1 Overview of Experiments 1 – 6**

| Materials[1] | Comprehension | Production | Population |
|---|---|---|---|
| Toys | **Chapter 4** | **Chapter 5** | |
| | 1a Act-out | 1b Elicitation | Dutch preschoolers |
| | 2a Act-out | 2b Elicitation | Dutch preschoolers |
| | | | |
| Animations | **Chapter 6** | **Chapter 7** | |
| | 3a Picture selection | 3b Elicitation | Dutch adults |
| | 4a Picture selection, Preferential looking | 4b Elicitation | Dutch preschoolers |
| | 5a Picture selection | 5b Elicitation | English adults |
| | 6a Picture selection, Preferential looking | 6b Elicitation | English preschoolers |

1. All experiments tested sentences with an animal-vehicle distinction except Experiment 2, which tested a human-vehicle distinction.

Chapter 4 and Chapter 5 report two aspects of the same experiments using toys. Chapter 4 presents the comprehension data from two act-out tasks with Dutch preschoolers, and Chapter 5 presents production data from two elicitation tasks carried out with the same populations. The first population of Dutch preschoolers was tested with sentences with an animal-vehicle animacy contrast and the second with sentences with a human-vehicle contrast. Chapter 6 and Chapter 7 also report two aspects of the same experiments, this time involving video animations and including measurements from automated eye tracking. Chapter 6 presents comprehension data from a picture selection task with Dutch- and English-speaking adults and preschoolers; the children were additionally tested with a preferential looking task. Chapter 7 presents production data from an elicitation task carried out with the same four populations tested in Chapter 6. All experiments reported in Chapter 6 and Chapter 7 involve only sentences with an animacy-vehicle distinction. Chapter 8 compares performance on the tasks across all of the experiments with regard to scorability and accuracy. Materials that were used appear in Appendix A and B, and individual results of preschoolers are listed in Appendix C.

In the final two chapters, the results of the experiments are discussed in light of the theoretical and methodological issues raised in Chapters 2 and 3. The discussion in Chapter 9 addresses where predictions were met, where predictions were not met, and possible explanations for discrepancies. Remarks are also made about the success of the tasks and materials used in the experiments. The dissertation is brought to a close by Chapter 10, in which conclusions are drawn about whether there is a developmental asymmetry between production and comprehension of word order and whether the relative animacy of subject and object has a systematic effect on sentence interpretation by preschoolers.

# 2 Accounting for asymmetry

## A theoretical model of production and comprehension

The traditional view of language is that comprehension is easier than production. It is common knowledge among linguists and parents alike that what children are able to understand can greatly exceed what they are able to produce. A well-known example of such an asymmetry is early phonological production: children understand words they are not yet able to properly produce themselves. For example, a young child who understands the word “cat” when uttered by an adult might say [ta] instead of [kæt] (Smolensky, 1996). The same holds for vocabulary development: young children say fewer words than they are able to understand (Goldin-Meadow, Seligman, & Gelman, 1976). The tendency for comprehension to precede production has also been found in the areas of morphology and syntax (Fraser et al., 1963).

Linguistic research has revealed additional, more subtle asymmetries in which production actually precedes comprehension. That is, children may consistently use a form correctly, but may not consistently interpret the same form correctly. This type of asymmetry has been observed with object pronouns (de Villiers et al., 2006), contrastive stress (Kuijper & Groothoff, 2010) and indefinite objects (Unsworth, 2007). An asymmetry in favor of production is more difficult to explain than one in favor of comprehension (Bates et al., 1995; Clark & Hecht, 1983; Ingram, 1974). It is, after all, counterintuitive that a child would consistently produce a form he or she cannot successfully interpret.

It is the aim of this research to determine whether early subject-object (S-O) word order is a further instance of production exceeding comprehension. In order to do so, I establish in this chapter which theoretical framework I use to make predictions about preschoolers’ and adults’ production and comprehension. These predictions are subsequently tested in experiments reported in Chapters 4 – 7.

The current chapter is broken down into four parts. First, I review previous research suggesting that English-speaking preschoolers can produce S-O word order more successfully than they can comprehend it. I begin with studies that have found variable comprehension and follow with studies that provide evidence for adult-like production. I pay particular attention to the results of Chapman and Miller (1975).

Second, I narrow down which theoretic approach most adequately accounts for production-comprehension asymmetries. For each acquisitional theory, I show how it accounts for imperfect comprehension and then assess how adequate it is after adult-like production is taken into consideration. Ultimately, I decide that the constraint-based framework Optimality Theory is the most appropriate grammatical model, able to simultaneously account for production and comprehension. Optimality Theory not only models early asymmetries and their disappearance, but it also allows specific predictions to be made about online processing, which is relevant to this research including real-time measurements of language use. I look specifically at the developmental model of Hendriks, de Hoop, and Lamers (2005) and the incremental, or real-time model of de Hoop and Lamers (2006).

Third, I describe the word order and animacy constraints used in these models. I explain the motivation behind adopting not only word order, but also animacy as a universal grammatical constraint. I do so by demonstrating that there are cross-linguistic preferences for word order and animacy in both comprehension and production. When discussing the animacy constraint, I address the crucial issue of why animacy is not expected to play a role in production. I illustrate how the ranking of these and other relevant constraints has been determined for adult-speakers of English, Dutch, and German.

Fourth, I suggest that children exposed to different types of input (i.e. different target languages) will come to rely on word order in comprehension at different rates. I look at variable comprehension of word order by preschoolers in languages other than English, focusing on a study by Chan, Lieven, and Tomasello (2009) in which English- and German-speaking preschoolers are compared. Under the assumption that German and Dutch provide similar input to children, I use Chan et al.'s results from German to make cross-linguistic predictions about Dutch preschoolers' word order comprehension. The chapter closes with a summary.

## 2.1 The asymmetry

What leads us to believe that there is asymmetry in how preschoolers comprehend and produce S-O word order? This section reviews the studies that provide evidence for (i) variable comprehension and (ii) adult-like production of word order by preschoolers. Taken together, the literature suggests that preschoolers are able to produce word order that they themselves cannot reliably understand. I conclude this section by asserting that this phenomenon can best be verified by means of a well-controlled experimental study.

### 2.1.1 Variable comprehension

It appears that children under the age of four fail to use word order in the same way as adults when interpreting semantically reversible sentences (English: de Villiers & de Villiers, 1973; Chapman & Miller, 1975; Chapman & Kohn, 1978; Thal & Flores, 2001; Chan, Meints, Lieven, & Tomasello, 2010; German: Lindner, 2003; Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008; Chan, Lieven, & Tomasello, 2009; Italian: Bates et al., 1984; Japanese: Hakuta, 1982). To what extent does the relative animacy of subject and object play a role in sentence comprehension? Focusing for the moment on English, I outline in this section first what is known about how children interpret pure word order, i.e. reversible sentences in which subject and object are equal (S=O) on the animacy hierarchy, like sentences (1) and (4). Then I look at studies that pit animacy against word order by testing children's interpretations of reversible sentences in which subject and object are not equal (S≠O) on the animacy hierarchy, like sentences (2) and (3).

| | | | |
|---|---|---|---|
| (1) | The boy is hitting the girl | [+an +an] | (animate-animate) |
| (2) | The girl is pulling the boat | [+an -an] | (animate-inanimate) |
| (3) | The truck is bumping the dog | [-an +an] | (inanimate-animate) |
| (4) | The car is pushing the truck | [-an -an] | (inanimate-inanimate) |

#### *2.1.1.1 Studies without animacy contrasts*

It is often said that children recognize the significance of S-O word order already at the one-word stage, with reference to the results of the first study to use the inter-modal preferential looking paradigm. Hirsh-Pasek and Golinkof (1996) tested 17-month-old English-learning infants to see if they paid attention to word order in reversible sentences. The paradigm was dubbed "inter-modal" because children were presented with both auditory and visual stimuli, as opposed to only auditory stimuli used in traditional preference paradigms common to phonological discrimination tasks. The children heard reversible S=O sentences such as *Big Bird is washing Cookie Monster* as they were simultaneously presented with two video screens. One screen showed the two animate characters performing the action described in the sentence; the other showed the reverse. Because the children preferred to watch the screen that matched the sentence they heard rather than the reverse, which was both semantically plausible and visually presented to them, it was interpreted that they must have been paying attention to word order.

Similar studies using the preferential looking paradigm tested whether English-speaking children preferred matching images over reversed images for sentences with novel verbs. Children can be tested on their comprehension of word order in sentences like *The duck is gorping the bunny* if they are first introduced to the new *gorping* action without having word order modeled (*Look, this is gorping!*). Gertner, Fisher, and Eisengart (2006) found that children at ages 1;9 and 2;1 looked longer to the matching screen after hearing sentences with novel verbs than would be expected by chance. Chan, Meints, Lieven, and Tomasello, (2010) tested children at ages 2;0, 2;9, and 3;5 on sentences with novel verbs (e.g. *gorp*, *meek*) and on sentences with familiar verbs (*push*, *kick*). They found that all of the children preferred the matching screen for sentences with familiar verbs, and the children at 2;9 and 3;5 looked longer to the matching screen for sentences with novel verbs. These studies together with the Hirsh-Pasek and Golinkoff (1996) study suggest that children use the order of words to determine agent- and patienthood when tested with tasks requiring only looking. In the case of the youngest children tested in these studies, they are interpreting word order even before they are able to combine words themselves (though not consistently for sentences with novel verbs). Because of the early age at which children seem to understand word order in S=O sentences, it is often assumed that comprehension precedes production of S-O word order, at least for English-speaking children.

When other tasks are used with S=O sentences, the picture is less clear. Some of the first studies that used picture selection tasks and act-out tasks were developed to test—among other constructions—word order in active, reversible sentences. In Fraser, Bellugi, and Brown's classic study (1963), twelve children between 3;0 and 3;6 were tested on the comprehension of ten different grammatical forms in English. For the word order test, each child received the sentence *the mommy is kissing the daddy* and *the train bumps the car* in a picture selection task. The child was presented with a picture of the correct action and a picture of the reversed action and was asked to select the correct picture. Of the 24 items administered (2 sentences over the twelve children), 16 received points to a picture reflecting an SO interpretation (66%). In a study involving a larger set of children with an act-out task, better rates of comprehension were found. Bever (1970; Bever, Mehler, & Valian, 1970) tested two groups of two- and three-year-olds (about 80 participants per age group) on the comprehension of word order in the following two sentences in an act out task: *the horse kisses the cow* and *the truck pushes the car*. The children were given two toys and asked to act out the sentences they heard. Bever found that actions reflecting an SO interpretation were given for the items nearly 100% of the time for each age group. Admittedly, the findings of these old but classic studies are not entirely in line with each other and their conclusions are based on item analyses rather than participant means. However, the results indicate children aged two and three have some knowledge of the significance of word order when interpreting S=O sentences. The studies were also an important first step in the development of effective methods of language assessment, which will be further discussed in Chapter 3.

**Table 2.1 Act-out studies testing word order comprehension of English-speaking children**

| Study | Sample sentence | Mean age | n | Mean proportion SO interpretations per sentence type[1] | | | |
|---|---|---|---|---|---|---|---|
| | | | | [+an +an] | [+an -an] | [-an +an] | [-an -an] |
| **Studies without animacy contrasts** | | | | | | | |
| de Villiers & de Villiers, 1973 | *Make the mommy kiss the daddy/ Make the boat bump the train* | 2;3 (MLU 2.25) | 7 | (.70) | - | - | (.70) |
| | | 2;3 (MLU 2.75) | 7 | (.81) | | | (.81) |
| | | 2;9 (MLU 3.25) | 4 | (.86) | | | (.86) |
| | | 2;9 (MLU 3.88) | 5 | (.83) | | | (.83) |
| Strohner & Nelson, 1974 | *The mouse pushes the bear* | 2;5 | 6 | .75 | - | - | - |
| | | 3;5 | 15 | .80 | | | |
| | | 4;3 | 15 | >.90 | | | |
| | | 5;3 | 15 | >.90 | | | |
| Bates, MacWhinney, Caselli, Devescovi, Natale, & Venza, 1984 | *The dog grabs the monkey* | 2;6 | 10 | .60 | na[1] | na | - |
| | | 3;6 | 10 | .88 | na | na | |
| | | 4;6 | 10 | .93 | na | na | |
| | | 5;6 | 10 | .95 | na | na | |
| Akhtar & Tomasello, 1997 (Exp.1) | *Make Mickey Mouse dack Ernie*[3] | 2;11 | 5 | .45 | - | - | - |
| | | 3;3 | 5 | .95 | | | |
| Akhtar & Tomasello, 1997 (Exp.2) | *Make Mickey Mouse dack Ernie* | 2;10 | 10 | .70 | - | - | - |
| | | 3;8 | 10 | .98 | | | |
| Thal & Flores, 2001[4] | *The bear is hitting the girl* | 2;0 | 21 | .52 | na | na | - |
| | | 2;6 | 23 | .70 | na | na | |
| | | 3;0 | 16 | mean across three types= .72 | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Chan, Meints, Lieven, & Tomasello, 2010 | *The duck is pushing the bunny/ The sheep is meeking the cat* | 2;0<br>2;9<br>3;5 | 23<br>22<br>22 | .39/.33<br>.82/.64<br>.99/.98 | - | - | - |
| **Studies with animacy contrasts** | | | | | | | |
| Chapman & Miller, 1975 | *The boy is pushing the car* | 2;2 | 15 | .67 | .94 | .50 | .65 |
| McClellan, Yewchuk, & Holdgrafer, 1986 | *The boy is pushing the car* | 2;8 | 10 | .49 | .11 | .76 | .41 |
| Childers & Tomasello, 2001 (Exp.2) | *The horse is dacking the truck* | 2;6 | 24 | .49 | .38 | .49 | - |
| Chan, Lieven, & Tomasello, 2009 | *The horse tams the telephone* | 2;6<br>3;6<br>4;6 | 25<br>24<br>25 | .78<br>.95<br>.99 | .86<br>.98<br>1.00 | .58<br>.97<br>.97 | - |

1. The proportions of SO interpretations reported were either calculated from all responses (including unscorable responses like unclear actions or a lack of a response) or from only scorable responses (SO interpretation vs. OS interpretation). In most cases, authors calculate proportions based on all responses and report that the number of unscorable responses was relatively low: De Villiers and de Villiers (1973: 335), Strohner and Nelson (1974: 569), Bates et al. (1984: 346), Childers and Tomasello (2001: 745), Chan et al. (2009: 283), Chan et al. (2010: 38), and Thal and Flores (2001: 181). In some cases, authors calculate proportions based on only scorable responses: Akhtar and Tomasello (1997: 957-958), Chapman and Miller (1975: 360), and McClellan, Yewchuk, and Holdgrafer (1986: 103).
2. If S≠O sentences were not adequately reversible, such as *The dog grabs the pencil*, I put "na" in place of the results. (The importance of controlling for verb selectional restrictions is discussed in Section 3.3.1.1.)
3. The sentences used in Akhtar and Tomasello (1997), Childers and Tomasello (2001), Chan et al. (2009), and Chan et al. (2010) contained novel verbs.
4. Because no significant influence of word order was found, percentages given for children of 2;0 from Thal & Flores (2001) are collapsed over canonical and noncanonical word order; therefore, scores include responses to sentences in the form noun-verb-noun, noun-noun-verb, and verb-noun-noun. For the same reason, the only percentage reported for children of 3;0 from this study was collapsed across the three word orders and the three sentence types.

The results of several later studies that used an act-out task to test word order comprehension in English-speaking preschoolers are summarized in Table 2.1. The table shows for each study a sample test sentence, the mean age of the group tested, the number of children from each age group, and their performance per sentence type.[1] The first seven studies listed provide information about how children interpret word order in reversible S=O sentences.

Upon hearing S=O sentences, children tested in the early study of de Villiers and de Villiers (1973) performed actions reflecting an SO interpretation at least 70% of the time—even children at age 2;3. (The results the authors provide collapse over [+an +an] and [-an -an].) They measured each child's productive language abilities on the basis of their mean length of utterance in morphemes (MLUm; Brown, 1973), which seemed only to make some difference for the younger age group, with longer MLUm linked to better comprehension. In later studies, Strohner and Nelson (1974) and Bates et al. (1984) found a steeper developmental pattern, with children interpreting word order as SO between 60% and 75% of the time at age 2;6, over 80% of the time at age 3;6, and over 90% of the time at ages 4 and 5. Skipping to Thal and Flores (2001), it was found that young two-year-olds did not significantly rely on word order when S=O (assuming that a *bear* and *girl* are equal in animacy, see Section 3.3.1.1.), with performance at 52% on the act-out task. The two-and-a-half- and three-year-olds, on the other hand, were able to rely somewhat on word order, with a success rate of about 70%.

Moving to studies run by Tomasello and colleagues, we see how children fare with novel verbs. In two experiments by Akhtar and Tomasello (1997) children under the age of three had some difficulty correctly responding to sentences like *Make Mickey Mouse dack Ernie*, with accuracy rates ranging from 45% to 70% accuracy rates, whereas children older than 3;3 had little difficulty (over 95% accuracy). The results of Chan et al. (2010) reported above suggested that children tested with the preferential looking paradigm could use word order well in sentences with familiar verbs and fairly well with novel verbs. The same set of children were additionally tested with an act-out task: at age 2;0 they were poor (39% with familiar verbs and 33% with novel verbs); the older two-year-olds performed better (82% familiar, 64% novel) and the three-year-olds the best (about 99% for both types of verbs).

These preferential looking and act-out studies investigated how successfully preschoolers understand the meaning carried by word order when tested with reversible sentences in which the subject and object are equal in animacy. Overall results suggest that English-speaking children aged one to three prefer to look at screens with events corresponding to the word order they hear. English-speaking children can also perform an action corresponding to the word order they hear, usually with at least 60% accuracy by about the age of 2;6 and at least 80% accuracy by the age of 3;6. Now I will look at the remaining studies in Table 2.1, which included an animacy manipulation, in order to see how it influences children's interpretation of S≠O sentences.

---

[1] Some studies include only some of the four sentence types listed in (1) – (4). Many also include extra factors, such as event probability, passive sentences, novel verbs, non-grammatical word orders, cross-linguistic data, or data from atypically developing children. Data in the table pertains only to typically developing English-speaking children. Furthermore, only data from fully reversible active sentences appear; hence, data is excluded for sentences that were in the passive voice or that had probable event or verb biases. When possible, data is also excluded for sentences with non-canonical word order, for example verb-noun-noun or noun-noun-verb.

#### *2.1.1.2 Studies with animacy contrasts*

Chapman and Miller (1975; henceforth C&M) wanted to determine what role animacy plays in the interpretation of word order by young English-speaking children. In order to do so, all four sentence types (1) – (4) were tested. That is, they included S=O sentences as well as S≠O sentences. The full list of materials used by C&M is provided in Table A.1 in Appendix A. Fifteen preschoolers between the ages of 1;8 and 2;8 were tested on comprehension with an act-out task. Three groups, each with five children, were defined on the basis of their mean length of utterance (MLUm).

Figure 2.1 shows the mean proportion of SO interpretations across sentence types in each of the three groups. The means collapsed over groups are shown in Table 2.1. On average, children performed roughly the same when acting out the two types of S=O sentences, with a mean of 66.5% for [+an +an] sentences and 65.2% correct for [-an -an] sentences. For [+an -an] sentences, children had SO interpretations 93.8% of the time on average: sentences like *The boy is pushing the car* were easy for the children to correctly act out. In contrast, [-an +an] sentences like *The car is pushing the boy* were interpreted as SO only 50.1% of the time on average. The children from the first two MLUm groups performed actions that reflected an SO interpretation less than half of the time, with rates of only 36.4% and 40.9%, respectively. These results are interpreted as evidence that children rely not only on word order, but also on the animacy of subject and object when determining the meaning of sentences.

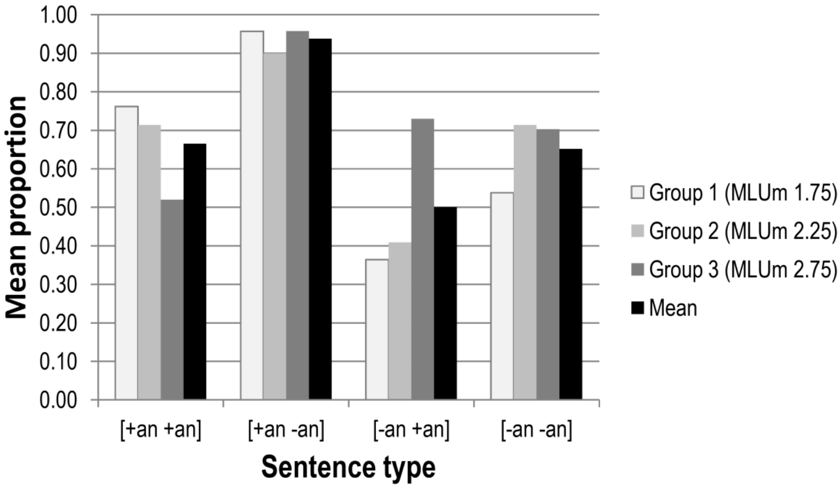


**Figure 2.1 Mean proportion SO interpretations per group and sentence type in Chapman and Miller (1975)**

In an attempt to replicate the findings of C&M, McClellan, Yewchuk, and Holdgrafer (1986) ran a similar act-out experiment with ten children aged about two-and-a-half. There were a few improvements in the design in that they tested more items per child and provided the child with only those toys needed for the test sentence at hand (rather than all six toys). Contrary to what C&M found, performance on the [+an -an] sentences was the lowest of all others, with a rate of only 11% SO interpretations. The authors suggest that their materials were not perfect, since some sentences differed in "semantic likelihood of occurrence" (McClellan, Yewchuk, & Holdgrafer, 1986: 110). They give the example that it is more likely that a car would hit a boy than that a boy would hit a car. (The importance of controlling for event probability is discussed in Section 3.3.1.2.) However, they used the same sentences as C&M, who found a very different pattern of results. The low rates of SO interpretations even for [+an +an] (49%) and [+an -an] sentences (41%) could be a sign that something else was awry in this study.

Subject and object animacy were also manipulated in a study by Childers and Tomasello (2001). They first demonstrated novel verbs like *dacking* to children aged 2;6, by showing two entities (in either a [+an +an], [+an -an], or [-an +an] situation). The experimenters performed the novel action and accompanied it with verbal input that indicated SO word order (*The horse is dacking the truck*), rather than the usual intransitive sentence used during novel verb demonstration (*This is dacking*). Similar to the results of McClellan et al. (1986), the [+an -an] sentences did not have an advantage over the [-an +an] sentences, with mean performance on both below 50%. Since the verbs that were used were novel, the results could not be due to biases of event probability. The overall poor performance could be an indication children at age 2;6 simply had difficulty in this task with novel verbs.

In a recent study by Chan, Lieven, & Tomasello (2009), children at ages 3;6 and 4;6 successfully interpreted sentences with novel verbs based on word order, with a rate of SO interpretations over 90% across the three sentence types. The children aged 2;6, on the other hand had SO interpretations of [+an -an] sentences quite often (86%) and for [+an +an] sentences almost as often (78%). In contrast, their performance on the [-an +an] sentences was low (58%). This pattern of variable comprehension for two-and-a-half-year-olds for novel verbs is in line with the findings of C&M for familiar verbs.

Having looked at studies both with and without animacy contrasts, we can now try to draw some conclusions about children's comprehension of word order. Despite the many differences between these studies with regard to theoretical goals, number of participants, additional conditions tested, etc., it can be concluded that comprehension of S-O word order is not consistent across sentence types (1) – (4). Rather, it is dependent on the relative animacy of the two entities in the sentence. Based on the discussion of the data presented in Table 2.1, the following pattern is construed for English-speaking preschoolers tested with an act-out task:

- At age 2;0, children do not seem to rely on word order or animacy information to interpret word order; their performance is often at 50%.

- Between the age of 2;6 and 3;0, children seem to make use of word order and animacy information simultaneously, with the following general performance ranges:

  S>O (85% – 95%) >> S=O (60% – 80%) >> S<O (50% – 60%)

- Between the age of 3;0 and 4;6, children come to rely on word order to interpret all active sentences; their performance is often between 75% and 100% overall.

Of course, this is a very rough pattern that excludes the anomalous findings of McClellan et al. (1986) and Childers and Tomasello (2001), two out of the four studies that included animacy contrasts. The fact that the children at age 2;6 in those two studies did not score above 50% on S=O sentences—sentences on which children in the other studies generally scored at least 60% indicates that these may not be reliable results. In any case, the present research aims to investigate whether this pattern holds in an experiment that is more carefully controlled than those listed in Table 2.1.

To summarize this section about variable comprehension of word order by young children: in preferential looking studies, English-speaking children at the one-word-stage seem to recognize the meaning conveyed by word order in sentences in which both entities are living. However, once the relative animacy of subject and object have been manipulated, preschooler's variable comprehension of word order is revealed.

### 2.1.2 Adult-like production

In contrast to preschoolers' variable comprehension of word order, their earliest productions conform to the word order rules of the adult language (Brown, 1973; Braine, 1976; Clahsen, 1982). Even when children first begin to combine words sometime between the ages of 18 and 24 months, "violations of normal [word] order are trifling few" (Brown, 1973: 156). According to a study of seven English-learning children (Ramer, 1976), the occurrence of word order errors ranged between 0% and 3.8%. Similar target production of word order has also been found in other languages such as German, Dutch, French, and Swedish, in which the ordering of constituents is affected by finiteness, negation, and scrambling (Verrips & Weissenborn, 1992; Poeppel & Wexler, 1993; Schaeffer, 2000; Neeleman & Weerman, 1997; Déprez & Pierce, 1993; Platzack, 1996).

Young children's apparent mastery of word order in their own productions was also found by C&M. The same group of children who were tested on comprehension in the act-out task with toys were also tested on production of the same types of sentences. The experimenter demonstrated an action with the toys, for example, by making the car hit the boy, and the children were asked to describe the action. The mean rate at which SO order was produced was compared to the mean rate at which word order was interpreted as SO. Results, listed in Table 2.2, show that SO order was used significantly more often in production (85.4%) than in comprehension (68.9%). Furthermore, mean rates of SO order used in production per sentence type, ranging between 82% and 89%, did not differ as an effect of animacy as they had in comprehension.

**Table 2.2 Mean percent SO responses in both tasks in Chapman and Miller (1975)**

| Sentence type | Comprehension | Production |
|---|---|---|
| + animate + animate | 66.5 | 83.7 |
| + animate - animate | 93.8 | 86.3 |
| - animate + animate | 50.1 | 89.3 |
| - animate - animate | 65.2 | 82.4 |
| Total | 68.9 | 85.4 |

According to C&M, the results of the two tasks indicate that "production precedes comprehension in grammatical acquisition for subject-object structure" (Chapman & Miller, 1975: 362). They find it puzzling why a child would not be able to "reverse his rule by inferring the relations among referent objects from the word order information" (1975: 369). Significantly better performance on production than comprehension of word order was also replicated in McClellan, Yewchuk, and Holdgrafer (1986: 108-110), with mean accuracy in production on each of the four sentence types ranging from 81% to 90%.

Another study, by Angiolillo and Goldin-Meadow (1982), had similar findings. They tested nine children between the ages of 2;4 and 2;11 on their production of sentences in which the animacy of subject and object were manipulated. The child watched as the transitive actions *hitting* and *pulling* were performed by animate or inanimate entities. Animate entities included adults such as the child's mother, father, and the experimenter and toys such as a boy doll, girl doll, elephant, and dog; inanimate entities included a toy tractor, airplane, tree, and cup. Angiolollo and Goldin-Meadow found that patients, whether animate or inanimate, were overwhelmingly placed post-verbally while agents, animate or inanimate, seldom occupied the post-verbal position. Eight out of the nine children said, for example, "the tractor's hitting the man" when the boy doll got hit, but "the man's hitting the elephant" when the boy doll was doing the hitting. Likewise, six out of the nine children could say, for the inanimate cup, "the boy hitting the mug" when the cup was hit, but "the mug hitting mommy hand" when the cup did the hitting. The authors conclude from this that the children in their study used word order to place agents before patients, regardless of animacy (Angiolillo & Goldin-Meadow, 1982: 631-2).

To sum up, the few studies that have directly tested the question of whether animacy affects word order in active sentences in production suggest that children adhere to word order in their own utterances in the face of animacy contrasts at the underlying, meaning level. This is in contrast to the variable word order found in the act-out studies testing comprehension discussed above. It seems, therefore, that children use word order better in their own productions than they are able to interpret it during comprehension.

What is needed to confirm this asymmetry is a well-controlled study that tests comprehension and production in the same set of children. It is not until the next chapter that I get into the details of the well-controlled design and materials of the current study, improving upon those used in the studies just outlined. First, a theoretical framework must be established that is able to make predictions about word order for both comprehension and production, for both adults and children, and for both English and Dutch.

## 2.2 The account

Before departing on the investigative journey of determining whether there is a word order asymmetry, the theoretical gear we take with us must first be established. In the first part of this section, several extra-grammatical and grammatical explanations of non-adult-like language in children are evaluated. Optimality Theory emerges as the most adequate framework to account for asymmetry because it incorporates into the grammar the separate demands of production and comprehension. In the second part of this section, the relevant Optimality Theoretic tools are expanded upon: Hendriks, de Hoop, and Lamers' (2005) model of early word order production and comprehension, and de Hoop and Lamers' (2006) model of grammar applied incrementally in real time. We will then be equipped with the basic concepts necessary for making predictions about how preschoolers and adults will perform in the current study.

### 2.2.1 Assessment of accounts

Variable performance by children, whether it be in comprehension or production, requires a theoretical explanation that addresses this non-adult-like behavior. For example, how in the case of object pronouns can it be accounted for that children mis-interpret the pronoun *him* yet correctly interpret the reflexive *himself*? Or why can preschoolers interpret word order in some sentences, but not others? What's more, when variable performance is found in only comprehension, but not in production (or vice versa), even more is required of the explanation. Why do children interpret reflexives better than pronouns, *and* why can they produce both in an adult-like way?

In the overview that follows, I give a concise assessment of six accounts of variable performance and asymmetries, making a distinction between (i) extra-grammatical and (ii) grammatical approaches. Extra-grammatical accounts blame sources of difficulty that lie outside the grammar, while grammatical explanations find fault within the grammar itself. For each account, I determine how adequately explanations of variable behavior can incorporate the additional phenomenon of asymmetry.

#### *2.2.1.1 Extra-grammatical accounts*

The following three extra-grammatical explanations assume that children have grammatical competence; but due to pragmatic, task, or processing limitations, children do not always show evidence of this knowledge.

*Pragmatic limitations*

According to the first extra-grammatical account, children have complete grammatical knowledge, but their non-adult-like comprehension is due to pragmatic limitations. The inability to incorporate information from the discourse level with information from the grammar obscures true knowledge. A distinction is made between a child's innate competence and learned pragmatic knowledge, with the latter as the culprit responsible for the child's mistakes. This type of explanation has been offered to explain children's variable comprehension of object pronouns (Chien & Wexler, 1990; Thornton & Wexler, 1999), and to the omission of pronouns in production (Weissenborn, 1992).

When considering an asymmetry between comprehension and production, Hendriks and Koster (2010) assert that this is not a viable explanation since it has yet to be explained why children would appear to have pragmatic knowledge in production, but not comprehension (or vice versa). For this reason it is unlikely that a pragmatic explanation can be adapted to a S-O word order asymmetry.

*Experimental limitations*

According to the second extra-grammatical account, children have complete grammatical knowledge, but their non-adult-like comprehension is due to experimental artifacts. The demands of the task(s) used to test comprehension obscure true knowledge (e.g. Crain, 1992: 373). It is often argued, for example, that experimental designs using sentence-picture verification tasks, often used with children over the age of four, do not meet the felicity conditions required to reject a sentence. This explanation is encouraged by research demonstrating that an alteration of this kind of task can considerably improve the performance of children (Conroy, Takahashi, Lidz, & Phillips, 2009; Crain et al., 1996; Davies & Katsos, 2010; Gualmini, 2004; Papafragou & Musolino, 2003). That children are put in unusual situations during a psycholinguistic test is also suggested by Keenan and MacWhinney (1987: 150), who believe that C&M "disabled normal comprehension" by removing contextual support that preschoolers are usually provided with in daily language use.

When considering an asymmetry between comprehension or production, this experimental limitation explanation continues to hold. After all, comprehension and production must necessarily be tested via different tasks. In fact, the results of C&M has been deemed by Bates, Dale, and Thal as one of the "few cases that dissociation actually seems to run in the opposite direction" than expected (1995: 8) and that "most researchers agree that [these] dissociations reflect cases in which [ . . . ] the comprehension test itself involves complex task demands that obscure the child's actual knowledge of grammatical structure." Basically, asymmetries under this view occur because one task obscures knowledge, whereas the other task does not.

Experimental limitations, however, may be too quickly appealed to (Hendriks & Koster, 2010; Tomasello, 2000). If across studies and tasks children show problems interpreting object pronouns, quantifiers, and scalar implicatures, it seems inappropriate to dismiss this collective performance as a byproduct of the tests themselves. Nevertheless, the explanation cannot be ruled out completely. Because of the possibility that asymmetry may result from methodology, different types of comprehension and production tasks, described in detail in Chapter 3, are used to test word order in the current study.

*Processing limitations*

According to the third extra-grammatical account, children have complete grammatical knowledge, but their non-adult-like comprehension is due to cognitive processing limitations. For example, C&M propose that the children show problems interpreting word order due to their inability to reverse cognitive processes (Chapman & Miller, 1975: 369). Reinhart (2006) attribute other instances of children's non-adult interpretations to limited processing capacity (e.g. limited working memory), which prevents children from using their underlying linguistic competence.

Good performance on production is already taken into account in C&M's processing explanation, in which children have difficulty calculating the meaning of a form, but do not have difficulty with the calculations necessary to produce the same form. In their own words: "the failure to use word cues in comprehension when word order is observed by the child in production is simply one instance of the preschooler's many failures to reverse processes that he can carry out in one direction" (1975: 369). However, if children cannot properly reverse the production process for successful comprehension, how is it that children did reliably interpret word order in at least some of the types of sentences tested? C&M believe children's success with S>O sentences results from an extra-grammatical *strategy* based on animacy information, but I will show below that there is an account that is able to incorporate use of animacy information into the grammar.

Reinhart (2004; 2006) attempts to account for good performance on production by claiming that ambiguity requires more cognitive processing in comprehension than in production. She believes that children have grammatical knowledge but exhibit imperfect comprehension of pronouns, contrastive stress, and scalar implicatures because more computation is required for comprehension than for production in these cases. Children lack the sufficient working memory necessary to reach an adult-like interpretation when faced with ambiguity because alternative derivations must be constructed and compared. Extra computation is not necessary in production since "semantically ambiguous derivations always pose a greater load on the hearer than on the producer (who always knows which meaning he or she intends)," (Reinhart 2004: 136).

This type of processing explanation underestimates, however, the amount of cognitive processing necessary for production. Production involves complex articulatory motor control, as well as a plethora of other processing computations like conceptual preparation, lexical selection, morphological encoding, etc. (Levelt, Roelofs, & Meyer, 1999). Speakers must also consider alternate derivations before producing an utterance, whether ambiguous or not (Hendriks & Spenader, 2004; Samek-Lodovici, 2007). Nevertheless, an extra-grammatical account appealing to cognitive processing limitations seems to remain valid for both types of asymmetries since one can argue that either comprehension or production of a certain construction requires more processing resources.

#### *2.2.1.2 Grammatical accounts*

In the following three types of grammatical explanations it is assumed that children do not yet have adult-like grammatical knowledge. Imperfect comprehension is due to a grammar that has not yet reached an adult-like state.

*Unset parameters*

According to the first grammatical account, children do not have grammatical knowledge necessary to prevent comprehension errors because they have not yet set the relevant parameter(s) that lead to adult-like interpretations (e.g. Guasti, 2002; Meisel, 1995). In the case of preschoolers' mistakes on the comprehension of S-O word order, it could be that the necessary parameters associated with constituent ordering are not yet correctly set (e.g. complement-head parameter and specifier-head parameters). Correct performance on some sentence types would then be attributed to an extra-grammatical strategy based on animacy information to aid in comprehension.

However, this parameter-setting explanation offered by nativists no longer holds once good performance on production is additionally considered. According to the parameter-setting theory, if a child's productions give evidence of grammatical knowledge of a construction, then the child must at the same time be able to comprehend the same construction. That is, if a child shows linguistic knowledge, then his or her parameter is set. Thus, a child's mistakes in comprehension can be explained by an unset parameter, but a child's additional good performance on production eliminates an unset parameter as a possibility.

*Ungeneralized rules*

According to the second grammatical account, children do not have generalized grammatical knowledge, but rather have knowledge limited to certain lexical items (e.g. verbs). Children first rely greatly on the specific input they receive and it is not until later that they are able to generalize the linguistic rules. Tomasello (1992) asserts that children learn grammatical relations like word order on a verb-by-verb basis. It is for this reason that many studies run by Tomasello and his colleagues involve novel verbs: if children have generalized knowledge of rules, they will be able to apply them to new verbs. In the case of preschoolers' mistakes on the comprehension of S-O word order, the young children in the C&M study would not always perform adult-like on the comprehension task since they have not yet learned how word order works with the familiar verbs being tested. Such an explanation would not further explain, however, why the children would do better on S>O sentences than the other sentence types.

This ungeneralized rules explanation offered by functionalists is also inadequate when good production is additionally considered—for the very same reason the parameter-setting explanation cannot handle asymmetry. If a child shows knowledge in production of word order this serves as proof that the child has the relevant linguistic knowledge. In the case of the children in C&M's study, the appropriate use of word order when producing sentences with familiar verbs like *hit* and *bump* suggests that they had already learned the transitive frames for at least those verbs. Yet the same children made mistakes in the comprehension of sentences using the very same verbs. Whether the knowledge they used in production was only verb-specific or already verb-general is not of importance. The problem with this explanation remains the same: children show knowledge of word order in production so it cannot be that they do not have this knowledge available to them in comprehension.

*Mis-ranked constraints*

Under this view, children have grammatical knowledge, but they do not yet use it in the same way as adults. According to Optimality Theory (OT), the grammars of children and adults are comprised of the same set of constraints, but certain constraints may be given greater priority in the grammar of the children than in the grammar of the adults. (Differences in constraint-ranking of a universal set of constraints are also believed to be the source of differences between languages.) This type of mis-ranking explanation has been offered for the mistakes that younger children make with phonological productions (Smolensky, 1996). There is also an OT account of the variable comprehension of word order in preschoolers; namely, that children place more importance on an animacy constraint than a word order constraint (Hendriks, 2008; Hendriks et al., 2005). Since constraints in OT are soft, or able to be violated, grammatical knowledge is not precluded just because the children in C&M allow non-adult-like interpretations or productions.

Similar to OT is the cue-based Competition Model (Bates & MacWhinney, 1982, 1987, 1989), a model of language processing in adults and children. The Competition Model asserts, like OT, that developmental (and cross-linguistic) differences are due to different weights given to *cues*, such as word order, agreement, or animacy information. Variable comprehension by children is due to their over-reliance on certain cues (Bates & MacWhinney, 1989: 58,65; Chan et al., 2009, 2010; Lieven & Ambridge, 2011: 231). However, the Competition Model is primarily a processing model of performance and does not attempt to offer a formal model of the native speaker's underlying grammatical knowledge, or competence, in the way that OT does (Bates & MacWhinney, 1989: 32-36).

OT is also able to account for asymmetry between comprehension and production within a grammatical framework. Constraints within this framework are characterized as "direction sensitive." That is, the effect a constraint has depends on whether a language user is speaking or listening, thereby allowing comprehension and production to be simultaneously modeled within a single grammar. Differences between comprehension and production are acknowledged by supporters of the Competition Model, (Bates & Devescovi, 1989: 229; Keenan & Macwhinney, 1987: 53), but there is no formalization of (a)symmetry. Thus, OT is best suited for accounting for asymmetry, whether comprehension exceeds production, as in the case of phonological productions (Smolensky, 1996), or production exceeds comprehension, as in the case of pronouns and early word order (Hendriks, 2008; Hendriks et al., 2005).

In sum, we have seen that there are extra-grammatical and grammatical ways to account for variable comprehension/production. Of the extra-grammatical explanations, the pragmatic limitations account was ruled out for a word order asymmetry. It was also pointed out that experimental artifacts and processing limitations are difficult to rule out since comprehension and production must be tested with different tasks and intrinsically involve different types of cognitive processing. When moving to grammatical accounts, both the nativists' parameter-setting approach and the functionalists' ungeneralized rules approach were ruled out due to their inability to account for asymmetry. Finally, Optimality Theory was presented as a constraint-based model of grammar that is able to account for both variable comprehension and developmental asymmetries, in a more formalized way than the cue-based Competition Model. I now explain more about how a constraint-based framework like OT works and present the mis-ranking model of early word order asymmetry.

### 2.2.2 Constraint-based account

Now that it has been established that a constraint-based account can best handle variable comprehension as well as asymmetry, I give here the details about precisely how a constraint-based system works. I first go over the basics of Optimality Theory, defining the concepts of input and output, "direction" of grammar, constraint hierarchies, tableaux, constraint "softness", and faithfulness versus markedness constraints. I explain the distinction between non-adult language caused by a failure to bidirectionalize and non-adult language caused by a mis-ranking of constraints. I go on to present Hendriks, de Hoop, and Lamers' (2005) grammatical model of word order asymmetry, which predicts that children will more easily interpret S>O sentences than S<O sentences due to a too highly ranked animacy constraint. An additional model is presented, namely de Hoop and Lamers' (2006) incremental model of grammar application in real time. Using their incremental approach, I am able to make predictions for how the four sentence types under discussion compare to each other with respect to processing difficulty.

#### *2.2.2.1 Optimality Theory*

Optimality Theory (Prince & Smolensky, 1993), models language use as a selection of the best *output* for a given *input*, based on a grammar comprised of constraints. The framework was originally proposed by Prince and Smolensky (1993; 2004) for modeling phonological processes, where the input is usually an underlying phonological form and the output is a surface form, or spoken word. OT has more recently been extended to model syntax and semantics: if the speaker's perspective is taken, as in OT Syntax, the input is a meaning and the output is a linguistic form, like a word or sentence (Bresnan, 2000; Grimshaw, 1995; Legendre, 2001); if the listener's perspective is taken, as in OT Semantics, the input is a form and the output a meaning, or interpretation (Hendriks & de Hoop, 2001). In other words, whether the output of the grammar is a form or a meaning depends on the *direction* in which the grammar is being used, i.e. production or comprehension.

The OT framework is unique in that it models grammar as a hierarchy of soft constraints, rather than as a set of hard rules (e.g. parameters). The set of constraints is believed to be (semi-)universal (see Fikkert and de Hoop, 2009 for a discussion), with individual languages differing only in their hierarchical ranking of the constraints. Tableaux are useful for illustrating the predicted outcome of a grammar with a certain constraint ranking. Take for example Tableau 2.1. Constraints appear at the top in an order that signifies their ranking; the more highly ranked a constraint, the farther to the left it appears. In this grammar, CONSTRAINT A is ranked more highly than CONSTRAINT B. Possible output candidates appear on the left. If a candidate violates a constraint, it receives a violation star under that constraint.

**Tableau 2.1 OT tableau**

| Input | CONSTRAINT A | CONSTRAINT B |
|---|---|---|
| Output candidate 1 | *! | |
| ☞ Output candidate 2 | | * |

In the example, Candidate 1 is ruled out due to its violation of CONSTRAINT A, (a "fatal" violation marked with "!"). Candidate 2 is the winning output (marked by a pointing finger) because it does not violate the most important constraint, CONSTRAINT A as Candidate 1 does; Candidate 2's violation of the less important CONSTRAINT B is thus tolerated. Another way to look at this system of soft rules is by using a non-linguistic situation. For example, if I want to decide whether to wear a warm but ugly sweater or a cute summer dress, and I have the highly ranked constraint DON'T FREEZE and the less highly ranked constraint STYLE, I will end up wearing the warm garment in the winter. Under this ranking, it is more important to stay alive, even at the cost of looking frumpy. Thus, in such a system, the optimal output among a set of possible candidates is that which *best* satisfies the total set of constraints and their ranking.

Another important aspect of OT is how the constraints determine the relationship between the input and the output. There are two types of constraints: *faithfulness* and *markedness*. Faithfulness constraints require that the relationship between input and output is as straightforward as possible. These constraints may have a similar effect when used in either comprehension or production. Markedness constraints, on the other hand, are only concerned with the output. A markedness constraint on meaning only helps distinguish between output meanings; likewise, a markedness constraint on form only helps distinguish between output forms. Markedness constraints are the source of differences between comprehension and production since, for example, a markedness constraint that affects comprehension (or output meanings only) simply has no effect in production.

The inherent asymmetry of grammar brought on by markedness constraints is what accounts for both variable comprehension/production and asymmetry in child language. While it is not agreed upon whether all constraints are innate (Kager, 1999; Prince & Smolensky, 1993) or if some are learned (Boersma, 1997; Fikkert & Levelt, 2004), it is often accepted that in the initial state of the grammar, markedness constraints outrank faithfulness constraints (M >> F; Boersma & Hayes, 2001; Legendre, Hagstrom, Vainikka, & Todorova, 2006; Tesar & Smolensky, 1998). Young children eventually correctly rank the constraints in their grammars by demoting markedness constraints (Tesar & Smolensky, 1998), promoting faithfulness constraints (Gnanadesikan, 2004), or both (Boersma & Hayes, 2001). Since markedness constraints are the source of asymmetry between production and comprehension, and young children have markedness constraints ranked too highly, it is possible that this incorrect ranking results in early asymmetries.[2] This type of mis-ranking (M >> F) has been offered as the cause of early phonological production delays by Smolensky (1996). He argues that children cannot accurately produce word forms that they themselves are able to understand because of a non-adult-like constraint ranking—with one or more markedness constraint on production being ranked too highly.

#### *2.2.2.2 Model of word order asymmetry*

Following Smolensky's (1996) account of phonological asymmetry, Hendriks, de Hoop, and Lamers (2005) attribute word order asymmetry found by C&M to the incorrect ranking of a markedness constraint above a faithfulness constraint. They demonstrate this by using two relevant constraints that help distinguish between subject and object, proposed by de Lamers and de Hoop (2004/2005): PRECEDENCE and PROMINENCE. The constraints are defined as follows:

> PRECEDENCE: The subject precedes the object
> PROMINENCE: The subject outranks the object in prominence (e.g. animacy)

PRECEDENCE requires the subject to precede the object and is formulated as a faithfulness constraint promoting a direct correspondence between the input and the output both in comprehension and production.[3] The second relevant constraint, PROMINENCE requires that the NP that is highest in animacy be interpreted as the subject and is formulated as a markedness constraint on output meaning only; as a consequence, PROMINENCE only plays a role during comprehension. This constraint is satisfied if S>O, violated once if S=O, and violated twice if S<O. De Hoop and Lamers (2006) show that adult speakers of English, Dutch, and German have the following ranking of these two constraints in their grammars: PRECEDENCE >> PROMINENCE. (These constraints and the motivation behind the adult ranking are discussed in detail in Section 2.3.)

---

[2] Early asymmetries in young children can be distinguished from later asymmetries (Hendriks & Spenader, 2005/6.; de Hoop & Krämer, 2005/6). Once children have ranked constraints, they must then learn to simultaneously consider both their own perspective and their conversational partner's perspective, i.e. bidirectional optimization (Blutner, 2000), as adults do. But before this ability is fully achieved, children between the ages of four and eight are said to exhibit asymmetry in favor of comprehension by not taking into account the hearer's perspective when speaking (overuse of subject anaphora: Wubs, Hendriks, Hoeks, & Koster, 2009), or they exhibit asymmetry in favor of production by not taking into account the speaker's perspective when listening (pronoun interpretation: Hendriks & Spenader, 2005/6; contrastive stress interpretation: Kuijper & Grothoff, 2010).

[3] Even though de Hoop and Lamers (2006) acknowledge that *subject* and *object* are grammatical labels that do not always link straightforwardly to the argument structure of a transitive verb, they choose to formulate PRECEDENCE and PROMINENCE using the terms *subject* and *object*. They intend the same relation I described in Figure 1.1 in Section 1.2.1.

Hendriks et al. (2005) demonstrate that a non-adult-like ranking of these constraints (PROMINENCE >> PRECEDENCE) could explain the variable comprehension of the English-speaking children in C&M's study. Tableau 2.2 shows how this ranking would adversely affect a child's comprehension of an S<O sentence like *The car is pushing the cow*. The first output meaning PUSH <car, cow> violates the highly ranked PROMINENCE constraint twice because the object outranks the subject in animacy. Even though the second output meaning PUSH <cow, car> violates PRECEDENCE since the first NP in the input form would then not be interpreted as the subject, this is still the candidate that best satisfies the constraints under this ranking. As a result, an OS interpretation is optimal for young children when interpreting a S<O sentence.

**Tableau 2.2 Children's comprehension of word order in S<O sentences (form to meaning)**

| "The car is pushing the cow" | PROMINENCE | PRECEDENCE |
|---|---|---|
| PUSH <car, cow> (SO interpretation) | **! | |
| ☞ PUSH <cow, car> (OS interpretation) | | * |

**Tableau 2.3 Children's production of word order in S<O sentences (meaning to form)**

| PUSH <car, cow> | PROMINENCE | PRECEDENCE |
|---|---|---|
| ☞ "The car is pushing the cow" (SO order) | ** | |
| "The cow is pushing the car" (OS order) | ** | *! |

The same incorrect ranking, however, would *not* adversely affect a child's production of an S<O sentence. Tableau 2.3 shows that a S<O meaning (the input) always violates PROMINENCE twice since the agent is inanimate and the patient is animate, regardless of the output form.[4] Since PROMINENCE cannot distinguish between the potential output forms, PRECEDENCE determines that only candidates in which the subject occurs before the object are optimal. In fact, SO order is expected for all four sentence types, illustrated in Tableau 2.4, since PROMINENCE simply does not distinguish between output forms in production. The predictions of this model are in line with the findings of C&M, and McClellan et al. (1986) whose preschoolers used SO order above 80% of the time when producing each sentence type.

The optimal meanings expected under a PROMINENCE >> PRECEDENCE ranking differ between sentence types in comprehension, shown in Tableau 2.5. As we already saw in Tableau 2.2, an SO interpretation of a [-an +an] sentence violates PROMINENCE, making an OS interpretation optimal. In contrast, an SO interpretation of a [+an -an] sentence is optimal because it satisfies PROMINENCE. Because both SO and OS interpretations violate PROMINENCE for [+an +an] and [-an -an] sentences, SO interpretations are ruled optimal by PRECEDENCE.

There are two ways in which the predictions of a grammar with ranking PROMINENCE >> PRECEDENCE do not precisely match the findings of the studies reviewed earlier in this chapter. First of all, the model shows that OS interpretations are optimal for S<O sentences, but children still had SO interpretations 50% – 60% of the time. Second, while the model predicts S>O, S=O >> S<O, the

[4] An alternative approach is to consider the constraint as never violated by any output form during production, i.e. empty cells across for all output form candidates. In either case (always a certain number of violations versus never any violations) the result is the same in that the constraint does not distinguish between output candidates during production.

following pattern was found in English-speaking preschoolers: S>O >> S=O >> S<O. That is, the model predicts that SO interpretations are optimal for both S>O and S=O sentences, but data shows a rate of about 60% – 80% for S=O that is generally not as high as the S>O sentences (85% – 95%).

**Tableau 2.4 Child's production of word order for four sentence types**

| Input meaning | Output form | PROMINENCE | PRECEDENCE |
|---|---|---|---|
| [+an +an] | ☞SO order | * | |
| | OS order | * | * |
| [+an -an] | ☞SO order | | |
| | OS order | | * |
| [-an +an] | ☞SO order | ** | |
| | OS order | ** | * |
| [-an -an] | ☞SO order | * | |
| | OS order | * | * |

**Tableau 2.5 Child's comprehension of word order for four sentence types**

| Input form | Output meaning | PROMINENCE | PRECEDENCE |
|---|---|---|---|
| [+an +an] | ☞SO interpretation | * | |
| | OS interpretation | * | * |
| [+an -an] | ☞SO interpretation | | |
| | OS interpretation | ** | * |
| [-an +an] | SO interpretation | ** | |
| | ☞OS interpretation | | * |
| [-an -an] | ☞SO interpretation | * | |
| | OS interpretation | * | * |

The re-ranking process itself may provide an explanation for the discrepancies. Remember that the child's grammar (PROMINENCE >> PRECEDENCE) must eventually come to look like that of the adult (PRECEDENCE >> PROMINENCE). Since early grammars are proposed to be in a process of gradual re-ranking (Boersma, 1997; Boersma & Hayes, 2001; Boersma & Levelt, 1999), it could be that a partial overlapping of the constraints in the hierarchy is responsible for the differences between the sentence types. That is, the children could benefit from the convergence of two constraints as in the case of S>O sentences, suffer from violations of higher ranked (though in principle "indecisive") constraints as in the case of S=O sentences, and benefit from a non-violated lower constraint, as in the case of S<O sentences. This type of explanation seems viable, especially when considering the differences between the children in C&M in Figure 2.1: the groups with lower MLUm did appear to have a preference for OS interpretations (about 60% or higher). The overall mean rate of SO interpretations was brought up by five children in the group with highest MLU, whose constraints conceivably started heading in the right direction.

In the end, the model of word order asymmetry is able to account for both variable comprehension and adult-like production. By placing the markedness constraint on comprehension that prefers animate subjects and inanimate objects above the faithfulness constraint that requires subjects to occur before objects, the grammar performs differently during comprehension than in production. While production is not affected by animacy, comprehension is, with the prediction that children will more likely correctly interpret S<O sentences than S>O sentences, with S=O somewhere in between due to possible constraint convergence. With this pattern established as a prediction for global, or sentence end interpretations, we turn now to real-time models of grammar to make predictions about behavior during sentence processing.

#### *2.2.2.3 Grammar as the parser*

Most language processing theories respect the competence-performance distinction. Remember, for example, that the Competition Model of language processing makes no claims about underlying competence, or that generativists believe that the mere use of language can obscure underlying knowledge. Clifton and colleagues (2003: 318) point out that experimental psychologists are "extremely cautious about basing cognitive processes on grammatical rules, which appear to change frequently with seemingly-arbitrary theoretical changes in linguistics." Consequently, competence and performance usually remain separate. Parsing strategies are believed to result from cognitive processing limitations, and they are not formulated as part of the grammar. Even if "strong competence" is assumed, the grammar still requires a few mechanisms for building and breaking down sentences, i.e. the parser (Ford, Bresnan, & Kaplan, 1982).

An alternative approach is to assume that the grammar is the parser, or at least that processing behavior can be defined in terms of the grammar. In recent years, this approach has been initiated (Phillips, 1995; Weinberg, 1993, 2001), largely within the OT framework (Artstein, 2000; Fanselow, Schlesewsky, Cavar, & Kliegl, 1999; Gibson & Broihier, 1993; Hoeks & Hendriks, 2009; Kuhn, 2000; Lamers & de Hoop, 2004, 2005; Stevenson & Smolensky, 2006; de Hoop & Lamers, 2006). The *incremental optimization* account set forth by Lamers and de Hoop (2004; 2005; de Hoop & Lamers, 2006) in particular uses grammatical constraints to model language comprehension in real time. Incremental optimization simply means determining the optimal output as each word is encountered in the input. There are two situations that Lamers and de Hoop look at to demonstrate how the grammar can predict processing behavior: switches in interpretation over time and qualitative differences in constraint violation patterns. The examples that follow all involve data in the form of event-related potentials (ERP), or the electric activity of the brain.

The first situation involves effects from a switch, or a jump from one optimal interpretation to another. De Hoop and Lamers (Lamers & de Hoop, 2005; de Hoop & Lamers, 2006) looked at ERP data that was collected as Dutch participants listened to transitive sentences that were either canonical (SO order) or object-fronted (OS order). As each word is encountered in sentence (5), an SO interpretation remains optimal. In contrast, an SO interpretation only remains optimal in sentence (6) until the pronoun is reached, at which point a violation of the case-marking constraint results in an optimal OS interpretation (nominative case-marked *hij* cannot be the object). De Hoop and Lamers note that ERP positivity, generally associated with syntactic violations or syntactically less preferred sentence continuations (Osterhout & Holcomb, 1992), found precisely at *hij* in (6). They posit that the positivity is a response to the "jump" from one optimal interpretation to another. Thus, grammatical constraint violations appear to have physiologically real effects during processing.

(5) *de oude vrouw* *verzorgde* *hem* (SO)
[the old lady]NOM/ACC nurse.PST him.3.M.SNG.ACC
'The old lady took care of him'

(6) *de oude vrouw* *verzorgde* ***hij*** (OS)
[the old lady]NOM/ACC nurse.PST he.3.M.SNG.NOM
'It is the old lady whom he took care of'

The second situation involves effects from qualitative differences, or differences in constraint violation patterns even when there is no switch in optimal interpretation over time. De Hoop and Lamers (2006) looked at data from a study of German discussed in Schlesewsky and Bornkessel (2004) that involved relative clauses like (7) and (8), shown in incremental Tableau 2.6. Because of the accusative case-marking on the first NP in both sentences, an OS interpretation remains optimal throughout the sentence fragment. The only difference between the pattern of violations is that PROMINENCE is violated once in clauses like (7) and twice in clauses like (8). That is, S<O is worse than S=O. According to de Hoop and Lamers, this difference in the severity of the PROMINENCE violation gives rise to the N400, or negativity associated with a semantic mismatch, that is found at the second NP in (8) but not in (7).

**Tableau 2.6 Incremental constraint violations of subordinate clauses in German**

| (7) | *welchen Bischof*<br>[which bishop]ACC<br>'which bishop | *der Priester*<br>[the priest]NOM<br>the priest | *begleitete*<br>acompany.PST<br>accompanied' |
|---|---|---|---|
| CASE | ✓ | ✓ | ✓ |
| PRECEDENCE | * | * | * |
| PROMINENCE | ✓ | * | * |
| ERP | | | |
| (8) | *welchen Bischof*<br>[which bishop]ACC<br>'which bishop | *der Zweig*<br>[the twig]NOM<br>the twig | *streifte*<br>brush.PST<br>brushed' |
| CASE | ✓ | ✓ | ✓ |
| PRECEDENCE | * | * | * |
| PROMINENCE | ✓ | ** | ** |
| ERP | | N400 | |

**Tableau 2.7 Incremental constraint violations of object relative clauses in English**

| (9) | The | child | that | the | revolver |
|---|---|---|---|---|---|
| PRECEDENCE | ✓ | ✓ | ✓ | * | * |
| PROMINENCE | ✓ | ✓ | ✓ | * | * |
| ERP | | | | | N400 |
| (10) | The | revolver | that | the | child |
| PRECEDENCE | ✓ | ✓ | ✓ | * | * |
| PROMINENCE | ✓ | * | * | ✓ | ✓ |
| ERP | | Neg.shift | Neg.shift | | |

Another example of a processing effect caused by a qualitative difference in violations of PROMINENCE comes from English. De Hoop and Lamers (2005) looked at data from Weckerly and Kutas (1999), who tested object relative clauses with either an animate or an inanimate antecedent, shown in incremental Tableau 2.7 with fragments (9) and (10). In both fragments, the same violation pattern of PRECEDENCE is found as the words are encountered since both contain object relative clauses.

Prominence, on the other hand has a different pattern for each type of fragment. When the subject of the relative clause is realized as inanimate, at the second NP, Prominence is violated in fragments like (9). De Hoop and Lamers propose that the N400 found at this point during processing reflects this grammatical violation. Interestingly, when participants encountered an initial inanimate NP (an early violation of Prominence), they exhibited negative shifts in brain activity that were not found when they encountered an initial animate NP (an early satisfaction of Prominence). This is interpreted as evidence that “the processing of an initial inanimate NP is more costly than an animate NP (De Hoop and Lamers, 2005: 167).

Using an incremental framework, I have modeled how sentence types (1) – (4) violate the constraints Precedence and Prominence over time during comprehension, shown in Tableau 2.8. In the adult grammar (Precedence >> Prominence), the global (i.e. end of sentence) prediction is always an SO interpretation without a switch; as each word is encountered in each type of sentence, Precedence is always the decisive constraint. However the pattern of Prominence violations are different between sentence types and also change over time. In the two sentence types with initial NPs that are animate, Prominence is not violated at any point, except in the case of [+an +an] sentences when the second animate NP is encountered. Following from the evidence just presented that initial inanimate NPs violate prominence, in the two sentence types with initial, inanimate NPs, Prominence is violated from the beginning. In the case of [-an +an] sentences, it receives a second violation when the second, animate NP is encountered. Thus, reversible, declarative sentences with inanimate subjects, especially S<O sentences, are predicted to be at a disadvantage during processing.

**Tableau 2.8 Incremental constraint violations of sentences with and without animacy contrasts**

| Input form | | Prec | Prom | | Prec | Prom | | Prec | Prom |
|---|---|---|---|---|---|---|---|---|---|
| [+an +an] | *$NP_1$* | ✓ | ✓ | *verb* | ✓ | ✓ | *$NP_2$* | ✓ | * |
| [+an -an] | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| [-an +an] | | ✓ | * | | ✓ | * | | ✓ | ** |
| [-an -an] | | ✓ | * | | ✓ | * | | ✓ | * |

If it is assumed that children process language in the same way as adults (cf. Clahsen & Felser, 2006), the predictions the incremental model makes about processing can be extended to developing grammars. What’s more, I propose that incremental patterns may affect the global, or sentence end outcome. That is, while adults are able to interpret sentences as SO in the face of subtle processing hindrances, children may allow them to affect their global interpretations. For example, a [-an -an] sentence (which has an early violation of Prominence) may be interpreted as SO less often by children than a [+an -an] sentence (which never violates Prominence), even though both should optimally receive a global SO interpretation by the end of the sentence. This is in line with earlier findings that children as old as seven commit themselves early to an interpretation of a sentence, which they fail to later revise (Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000: 623; Sekerina, Stromswold, & Hestvik, 2004: 125; Trueswell, Sekerina, Hill, & Logrip, 1999: 19). This would also explain why children seem to understand S>O sentences more easily than S=O sentences.

To summarize this section describing the constraint-based model of an early word order asymmetry: Optimality Theory is a system of ranked, soft constraints. Because of the nature of markedness constraints, it makes a distinction between how grammar functions during comprehension and production. Hendriks et al. (2005) propose that an early word order asymmetry stems from an incorrect ranking of constraints in the grammar of preschoolers, namely that an animacy constraint on meaning is ranked above a word order constraint on form and meaning. A constraint-based grammar can also be used to make predictions about how easy the four sentence types under discussion are to process, relative to each other. These two models back up the pattern established in the previous section: S>O >> S=O >> S<O. In the next section, constraints take the center stage as I review how the word order and animacy constraints have been motivated and ranked in the OT literature.

## 2.3 The constraints

The conversation has already begun to extend itself from English to Dutch and German when examples of incremental processing in OT were presented. In this section, I first review the cross-linguistic evidence that motivates treating word order and animacy as universal grammatical constraints. I clarify why the animacy constraint is considered to be a markedness constraint on meaning only. I then present the motivation for the adult ranking of the distinguishability constraints in German, Dutch, and English.

### 2.3.1 Motivation of constraints

Of the two constraints PROMINENCE and PRECEDENCE, there is a stronger intuition that word order is a linguistic universal. Word order is considered to function as a part of syntax—and therefore as a part of the grammar. Animacy, on the other hand, is considered in some frameworks to be a factor that influences comprehension and production, but this contribution is usually seen as lexical rather than grammatical. I explain the motivation behind treating both word order and animacy as a universal constraint.[5] I first very briefly present evidence that has been used to motivate Precedence as a universal constraint on production and comprehension, before giving evidence that has been used to motivate PROMINENCE as a universal constraint on comprehension only.

#### *2.3.1.1 Universality of word order*

What is evidence that PRECEDENCE is a universal constraint? Beginning with production, strong evidence of the preference for the subject to precede the object comes from typology. In a sample from Dryer (2008), there are 1018 languages with basic SOV, SVO, or VSO word order and only 39 with basic VOS, OVS, or OSV word order. If the additional 171 languages with no dominant word order are taken into account, then SO word order is the basic word order of 83% of the world's languages—a clear majority.

Regarding comprehension, there is psycholinguistic evidence from English that the first NP is preferably interpreted as the subject. For example, subject relative clauses (*the babysitter that chased the child*) are easier to interpret than object relative clauses (*the babysitter that the child chased*) (Traxler, Morris, & Seely, 2002, Experiment 1). The same preference has been found to hold in German and Dutch (cf. Kaan, 1997 for review) for relative clauses (Frazier, 1987; Mecklinger, Schriefers, Steinhauer, & Friederici, 1995; Schriefers, Friederici, & Kühn, 1995) and declarative sentences (Frazier & Flores D'Arcais, 1989; Knoeferle & Crocker, 2005; Lamers, 2005; Weber, Grice, & Crocker, 2006).

---

[5] An argument has been levered against Optimality Theoretic syntax and semantics regarding the apparent ease with which constraints are functionally motivated. According to Newmeyer (2002) the "plausible external motivations are so numerous, so diverse, and so open-ended that any conceivable rule or constraint in any framework could be provided with one" (Newmeyer, 2002: 56). This "caricature" of the OT framework, however, is refuted by Bresnan and Aissen (2002), who point out that constraints are supported by the detailed work of linguists in many fields, such as descriptive fieldwork, corpus studies, typology, computational linguistics, and psycholinguistic experiments (Bresnan & Aissen, 2002: 85). In this spirit, I present evidence from different areas of linguistic research that has been used to motivate Precedence and Prominence.

#### *2.3.1.2 Universality of animacy*

According to many linguists, the influence of animacy on comprehension in English is recognized, but not regarded as a systematic part of the grammar (e.g. MacWhinney, Bates, and Kliegl, 1984: 137, see citation in Section 1.1.2.2). The fact that animacy plays a more absolute role in other languages, however, is evidence that there is a universal animacy constraint in human grammar, according to principles of OT. I present here examples demonstrating that animacy operates overtly in lesser known languages, followed by evidence that there is also a systematic influence of animacy in English and other Germanic languages. I then make some remarks about how PROMINENCE is formulated, and why it is not expected to influence production.

In lesser known languages, violation of a constraint requiring the NP highest in animacy to be the agent results in an inappropriate interpretation. The following example of Awtuw (Feldman, 1986) is discussed by de Swart (2007). In Awtuw, an indigenous language of Papua New Guinea, the meaning BITE <pig, lady> would not be the optimal interpretation of sentence (11). This is because the NP that is highest in animacy is to be automatically interpreted as the agent. In order to indicate that the NP that is highest in animacy (the lady) is the patient, the speaker of Awtuw must give it object marking, as in (12).

| | | | | | |
|---|---|---|---|---|---|
| (11) | *Tey* | *tale* | *yaw* | *d-œl-I* | (SO) |
| | 3.F.SG | woman | pig | FAC.bite.PST | |
| | 'The woman bit the pig' | | | | |

| | | | | | |
|---|---|---|---|---|---|
| (12) | *Tey* | *tale-re* | *yaw* | *d-œl-I* | (OS) |
| | 3.F.SG | woman.OBJ | pig | FAC.bite.PST | |
| | 'The pig bit the woman' | | | | |

Likewise in Fore (Scott, 1987), another Papuan language discussed by de Swart, the NP that is highest in animacy is automatically interpreted as the agent. In Fore, however, the agent rather than the patient is marked in order to indicate a non-default meaning.

The preference for an animate subject in comprehension is not bound to these lesser known languages. There is psycholinguistic evidence from English that suggests the NP highest in animacy is preferably interpreted as the subject. For example, relative clauses are easier to interpret when the subject of the relative clause is animate and the object of the relative clause is inanimate (Clifton et al., 2003; Traxler et al., 2002: Experiment 3; Trueswell & Tanenhaus, 1994; Weckerly & Kutas, 1999). In other words, the normal difficulty associated with object relative clauses can be reduced when the clausal patient is inanimate. The same effect has been found by Mak, Vonk, and Schriefers (2002) for Dutch.

Because of the convincing cross-linguistic evidence that there is a universal constraint on animacy when interpreting sentences, proponents of OT have formulated such a constraint, of which there are a few versions:

BIAS (Zeevat & Jäger, 2002) prefer a reading that is available in most cases

BIAS (Jäger, 2003) an NP of a certain morphological category is interpreted as having the grammatical function that is most probable for this category

BIAS (Aissen, 2004): a sentence interpretation must satisfy the following requirements

Subjects are 1st/2nd person, human, definite, high topichood
Objects are 3rd person, inanimate, indefinite, low topichood

PROMINENCE (Lamers & de Hoop, 2004/2005): the subject outranks the object in prominence (specificity, definiteness, saliency, animacy, and pronominal person—perhaps also topichood and givenness)

BIAS (de Swart, 2007): interpret a sentence according to the following regularities
Transitive Subject (A): +animate, +definite, given, pronominal, topic
Direct Object (O): ± animate, ±definite, new, nominal, comment

All versions presuppose that hearers make use of statistical regularities during interpretation—with the first version by Zeevat and Jäger being the least specific about what those regularities are. Later formulations more specifically define prototypicality and do so with reference to properties of NPs such as definiteness, givenness, and animacy. De Swart is more restrictive about what is prototypically a subject (e.g. should be animate) than what is prototypically an object (e.g. can be animate or inanimate). In contrast, Lamers and de Hoop, as well as Aissen, treat subject and object as more complementary, highlighting the importance of the relative difference between subject and object with regard to a property like animacy.

What these animacy constraints have in common is that they capture a universal preference among language users for a certain type of interpretation, or output meaning. An animacy constraint like PROMINENCE is a markedness constraint that plays a role in comprehension, but it has no effect in production. Crucially, *a preference for a certain meaning cannot distinguish between the types of forms that will be used to express that type of meaning*. This aspect of OT markedness constraints is central to Hendriks et al.'s model of the word order asymmetry.

*Animacy in production*

The existence of a constraint that requires a prominent NP to be interpreted as subject during comprehension does not preclude a preference to give a prominent NP early placement or high grammatical function during production. On the contrary, prominence is known to also influence production. Below, I show how animacy appears to affect production and give reasons for why animacy is nevertheless not expected to affect production in an experiment like that of C&M.

How does animacy affect the production of transitive sentences? Animacy appears to affect word order (Kempen & Hoenkamp, 1987; Tomlin: 102-139, 1986; de Smedt, 1990): speakers of English, Dutch, and German tend to place animate NPs before inanimate NPs in sentences (Bock & Loebell, 1990; Bouma, 2008: 256; Ferreira, 1994; McDonald, Bock, & Kelly, 1993; Prat-Sala & Branigan, 2000; Van Nice & Dietrich, 2003; van Bergen, 2011). The same tendency was found in two- to eight-year-old speakers of English (Clark, 1965; Dewart, 1979; Lempert, 1989). In addition, in some languages the most animate NP of a transitive sentence must come first, forcing a passive construction if the patient is more animate than the agent: Navajo (Native American: Comrie, 1989: 193) and Tzotzil (Mayan: Aissen, 1997).

Animacy also appears to affect grammatical function assignment. Regardless of order, higher animacy equates to higher grammatical functions, using a scale such as subjects >> direct objects >> indirect objects (Bock & Warren, 1985; Bock, Loebell, & Morey, 1992; McDonald et al., 1993). Corpus studies of Germanic languages show that in transitive constructions, animate NPs usually occur as subjects and inanimate NPs usually occur as objects (Dutch: Bouma, 2008; Swedish: Dahl, 2000; Dahl & Fraurud, 1996; German: Kempen & Harbusch, 2004; English: McDonald, 1987; Norwegian: Øvrelid, 2004). The result is frequent sentences with subjects that are higher in animacy than objects, e.g. 86% S>O, 13% S=O, and less than 1% S<O sentences in a corpus of Dutch adult speech (Bouma 2008: 257), and 95% S>O, 5% S=O, and 0% S<O in Dutch child speech (Hogeweg & de Hoop, 2010). In addition, in some languages an inanimate transitive subject is ungrammatical, such as Jacaltek (Mayan), Lakhota (Native American), and Japanese (cf. de Swart, 2007).

Based on these findings, it seems that prominence also affects production. However, this effect has been defined primarily in terms of motivations from the discourse level. For example, an object may be placed before the subject by a Dutch speaker if he knows that the hearer will be able to determine grammatical function using sources of information other than word order, like information from the discourse concerning topichood or givenness (Bouma, 2008: 234). Prominence can also be seen as a property that determines how accessible, or easily retrieved an NP is by speakers during production (Branigan, Pickering, & Tanaka, 2008; Tanaka, Branigan, & Pickering, 2005; van Bergen, 2011: 20-25). Information that is the topic of conversation or given in the discourse has high *contextual accessibility*, so such prominent information may receive early placement or a higher grammatical function in a sentence.

However, animacy stands apart from the other sorts of prominence in that it can be defined independently of context. While definiteness or givenness are determined by the discourse, animacy alone is an inherent lexical feature that cannot be changed by the discourse or altered by morphological marking (Prat-Sala & Branigan, 2000; de Swart, 2007: 135-195). A distinction has therefore been made between derived, *contextual accessibility* (e.g., definiteness, givenness) and conceptual, or *inherent accessibility* (e.g. animacy). Branigan, Pickering, and Tanaka (2008: 175) point out that animate entities are more likely to be worthy of discussion, hence more likely to be talked about at length, hence more likely to be given/salient in a conversation. As a result, animate entities are likely to have high contextual accessibility in addition to the pre-existing inherent accessibility.

It would appear then that prominence largely affects production for discourse reasons: the speaker can stray from default word order if the context allows it, and accessibility might drive early placement of contextually prominent NPs. *Inherent prominence* (animacy) alone is not expected to play a significant role in the production of word order without additional *contextual prominence* derived from the discourse. In comprehension, on the other hand, inherent prominence is believed to independently affect how sentences are interpreted. In other words, I am suggesting that word order alternations in production are primarily motivated by contextual reasons rather than by mere differences in animacy between entities. Thus, word order is not expected to be affected by animacy when produced out of context in an experimental setting like that of C&M.

### 2.3.2 Motivation of adult ranking

The constraints PRECEDENCE and PROMINENCE discussed so far are only two of the five universal DISTINGUISHABILITY constraints that de Hoop and Lamers (2006) have proposed. The following constraints distinguish which argument in a transitive sentence is the subject and which is the object, two of which we have already seen:

CASE: the subject is in the nominative case, the object in the accusative
AGREEMENT: the verb agrees with the subject
SELECTION: fit the selectional restrictions of the verb

PRECEDENCE: the subject (linearly) precedes the object
PROMINENCE: the subject outranks the object in prominence (e.g. animacy)

The ranking of PRECEDENCE in a particular language relative to the other constraints will reflect how flexibly word order functions in that language. As is shown below, the strict word order of English is modeled by ranking PRECEDENCE high in its grammar relative to the other constraints. Dutch and German, on the other hand, allow other constraints to play a more important role than word order. All three languages, however, rank PRECEDENCE more highly than PROMINENCE.

English differs from Dutch and German in its inflexibility of word order: While all three languages allow OS word order for communicative functions like topicalization, question-answering, or for emphasis or contrast, such occurrences are rare in English (Snider & Zaenen, 2006: 327). Sentences (13) and (14) are allowed in English (Bates et al., 1984: 344; MacWhinney, Bates, & R Kliegl, 1984: 132), though (15) is not.

| | | |
|---|---|---|
| (13) | Egg creams, I like | (OSV) |
| (14) | Makes a mean apple pie, my old lady | (VOS) |
| (15) | *Juice want I | (OVS) |

Dutch and German, on the other hand, are facilitated by verb-second (V2) syntax.[6] and allow OS order more freely. Though OVS order is allowed in Dutch, a corpus study by Bouma shows that object-fronting occurs in only about 14% of Dutch spoken sentences with a direct object (2008).

The fact that word order is somewhat flexible in Dutch and German is reflected in de Hoop and Lamers' (2006) proposed ranking of the DISTINGUISHABILITY constraints for the languages, which places several constraints above PRECEDENCE:

{CASE, AGREEMENT} >> SELECTION >> PRECEDENCE >> PROMINENCE

They motivate this ranking by using the optimal/preferred interpretations of German sentences (20) – (27) in Tableau 2.9. An SO interpretation of sentence (20) violates none of the constraints. Sentences (21), (22), (23) and (25) demonstrate that an interpretation that violates CASE or AGREEMENT is severe, and is

---

[6] As verb-second (V2) languages, Dutch and German require that the finite verb take position as the second constituent of matrix clauses. The first position may be filled by a subject as in (16) and (18), adverbial or prepositional phrase as in (17), or an object (19).

(16) *Ik* *wil* *sap* *drinken* ($SV_{FIN}OV$)
I want.FIN juice drink.INF
'I want to drink juice'

(17) *Met* *plezier* *heb* *ik* *sap* *gedronken* ($V_{FIN}SOV$)
with pleasure have.FIN I juice drink.PST
'I drank juice with pleasure'

(18) *Ik* *drink* *sap* ($SV_{FIN}O$)
I drink.FIN juice
'I am drinking juice'

(19) *Sap* *wil* *ik* *drinken* ($OV_{FIN}SV$)
Juice want.FIN I drink.INF
'It's juice that I want to drink'

rejected in favor of interpretations that do not violate these constraints, even if it means violating SELECTION, PRECEDENCE, or PROMINENCE. Case is marked in German on the determiner of the noun phrase (and attributive adjectives), in these examples as either nominative or accusative. Agreement is marked on the verb, in these examples as either singular or plural, and 1st person or 3rd person. Sentences (24) and (26) demonstrate that an interpretation that violates SELECTION is more severe than either a violation of PRECEDENCE or PROMINENCE. The final sentence (27) shows that violations of PRECEDENCE are more severe than PROMINENCE. (These are precisely the types of sentences that we expect children to interpret as OS, based on studies discussed in Section 2.1.1.2.) Although Dutch does not have case-inflection on NPs, case-inflected pronouns may cause an OS interpretation of sentences like (6), thus de Hoop and Lamers assume that Dutch has the same ranking as German.

As for English, it is assumed by de Hoop and Lamers that PRECEDENCE at least outranks PROMINENCE. Since the present discussion is concerned primarily with sentences like (27) in which agreement, case, and selection are kept constant, this ranking of only the subset PRECEDENCE and PROMINENCE will suffice. But we can go a bit further using intuitions about sentences like *The hamburger is eating the boy* in Chapter 1, which suggest that in English, PRECEDENCE likely outranks SELECTION as well. An ungrammatical, but interpretable sentence like *Me are hitting the boys*, which is likely to be interpreted as SO despite violations of CASE and AGREEMENT, suggests that PRECEDENCE is the most highly ranked of the five DISTINGUISHABILITY constraints in English. This results in the following (partial) ranking for English:

PRECEDENCE >> {CASE, AGREEMENT, SELECTION, PROMINENCE}

These suggested rankings are in line with other rankings offered for English, Dutch, and German by Bates and MacWhinney (1989:44). They obtained the following *hierarchies of cue strength* by testing adults across the different languages on their agent-patient assignment in sentence comprehension:

| | |
|---|---|
| English: | Word Order > Animacy, Agreement |
| Dutch: | Case > Word Order > Animacy |
| German: | Case > Agreement > Animacy > Word Order |

A notable difference is that they found animacy to be more important in determining meaning than word order in adult German. This is probably due to the fact they fail to include a distinction between the selection properties of the verb and the animacy properties of the NPs. While the two types of information are related, Hoop and Lamers have shown that they are separate. Sentences with verb restrictions like (24) and (26) are not interpreted on the basis of word order, whereas sentences like (27) in which the verb is neutral *are* interpreted on the basis of word order—despite the difference in animacy of the NPs. The difference in preferred interpretations of these types of sentences merit a crucial distinction between relative NP animacy and verb selection restrictions.

To summarize this section focusing on constraints: we have seen why not only word order, but also animacy is treated as a universal constraint. Furthermore, it has been discussed why PROMINENCE is not considered to play a role in the production of sentences like those tested in C&M. Finally, the motivation behind the proposed constraint rankings for adult speakers of German, Dutch, and English was discussed, touching on differences that make word order in German and Dutch somewhat more flexible

**Tableau 2.9 Motivation of German ranking given by de Hoop & Lamers (2006)**

| | | Interpretation | CASE | AGR | SEL | PREC | PROM |
|---|---|---|---|---|---|---|---|
| (20) | *Ich* *habe* *den Zaun* *zerbrochen*<br>I.NOM have.1.SNG [the fence]ACC break.PTC<br>'I broke the fence' | ☞SO | | | | | |
| | | OS | * | * | * | * | * |
| (21) | *Den Zaun* *habe* *ich* *zerbrochen*<br>[the fence]ACC have.1.SNG I.NOM break.PTC<br>'I broke the fence' | SO | * | * | * | | * |
| | | ☞OS | | | | * | |
| (22) | *Den Lehrer* *hat* *der Jungen* *geschlagen*<br>[the teacher]ACC have.3.SNG [the boy]NOM hit.PTC<br>'The boy hit the teacher' | SO | * | | | | * |
| | | ☞OS | | | | * | * |
| (23) | *Die Lehrerinnen* *hat* *Bernard* *geschlagen*<br>[the teachers]NOM/ACC have.3.SNG Bernard.NOM/ACC hit.PTC<br>'Bernard hit the teachers' | SO | | * | | | * |
| | | ☞OS | | | | * | * |
| (24) | *Die Jacke* *hat* *Bernard* *gesehen*<br>[the coat]NOM/ACC have.3.SNG Bernard.NOM/ACC see.PTC<br>'Bernard saw the coat' | SO | | | * | | * |
| | | ☞OS | | | | * | |
| (25) | *Der Zaun* *hat* *Bernhard* *zerbrochen*<br>[the fence]NOM have.3.SNG Bernard.NOM/ACC break.PTC<br>'The fence broke Bernard' | ☞SO | | | * | | * |
| | | OS | * | | | * | |
| (26) | *Bernard* *hat* *die Vorstelling* *deprimiert*<br>Bernard.NOM/ACC have.3SNG [the performance] NOM/ACC depress.PTC<br>'The performance depressed Bernard' | SO | | | * | | |
| | | ☞OS | | | | * | * |
| (27) | *Die Pflanze* *streifte* *Bernard*<br>[the plant]NOM/ACC brush.SING.PST Bernard.NOM/ACC<br>'The plant brushed Bernard' | ☞SO | | | | | * |
| | | OS | | | | * | |

than in English. I now look at how this cross-linguistic difference predicts that Dutch-speaking preschoolers will take longer to come to rely on word order than English-speaking preschoolers.

## 2.4 The role of input

Hendriks et al.'s model uses data from English to model early word order comprehension, but they also assume the model holds for early word order in both English and Dutch. In the previous section, we saw that there is a relatively lower ranking of word order in German and Dutch compared to English. How might this affect how German- and Dutch-speaking children comprehend word order compared to English-speaking children? By revisiting results of the study by Chan, Lieven, and Tomasello (2009), this time including their results from German-speaking preschoolers, I am able to make cross-linguistic predictions for Dutch-speaking preschoolers.

### 2.4.1 Variable comprehension beyond English

So far we have only looked at studies carried out with English-speaking preschoolers. As was the case with the review of studies in English in Section 2.1.1, much of the data from studies of other languages was collected using sentences that confounded verb selectional restrictions with animacy. For instance, a pair of studies by Bates and MacWhinney and their colleagues compare German, English, and Italian children's and adults' comprehension of S≠O sentences with verbs like *eat*, *bite* and *grab* together with inanimate nouns like *rock*, *ball*, and *pencil* (Bates et al., 1984: 345; MacWhinney, Bates, & Kliegl, 1984: 139)

There are some reliable studies of languages other than English that suggest that animacy does serve as an important early source of information when interpreting transitive sentences. An investigation of children acquiring Warlpiri, an indigenous Australian language, revealed that three-year-olds first rely on animacy, rather than adult-preferred case marking, when interpreting transitive constructions (Bates & MacWhinney: 45, 1989; Bavin & Shopen, 1989:195). In a study by Lindner (2003) German children between the age of two and three relied more on animacy than word order, case, or agreement (reliable results if we assume that a block in the context of an act-out task is an acceptable agent for *push*).

One study in particular allows us to directly compare the behavior of children learning a strict word order language like English with that of children learning a more flexible language like German. The results of Chan, Lieven, and Tomasello (2009) were presented already in Section 2.1.1.2 as a study that tested English-speaking children on both S=O and S≠O sentences. Children were asked to act out sentences with novel verbs like *The horse tams the telephone.* Remember that the results were in line with that of C&M, showing an S>O >> S=O >> S<O pattern for children at two-and-a-half, a pattern that was no longer present in the children they tested at age three-and-a-half or four-and-a-half. Within the same study, Chan et al. (2009) also tested German-speaking children. The results from English-speaking children in Table 2.1 are presented again here in Table 2.5, alongside results from the German children. Three age groups of children in each language were tested with each group containing about 25 participants. German sentences were neutral with regard to case- and agreement-marking.

Cross-linguistic results show that the English-speaking children mastered word order more quickly than the German-speaking children. To begin with overall development, there was an improvement with age on each sentence type for each language from age 2;6 to 3;6, but only in the case of the German children was there an improvement again on each sentence type from age 3;6 to 4;6. The English-speaking children did not improve from age 3;6 to 4;6 because they were already at ceiling by the age of 3;6.

**Table 2.5 Mean proportion SO interpretations by German- and English-speaking children from Chan, Lieven, and Tomasello (2009)**

| | English Age group | | | German Age group | | |
|---|---|---|---|---|---|---|
| Sentence type | 2;6 | 3;6 | 4;6 | 2;6 | 3;6 | 4;6 |
| [+an -an] S>O | .86 | .98 | 1.00 | .71 | .88 | .96 |
| [+an +an] S=O | .78 | .95 | .99 | .67 | .89 | .99 |
| [-an +an] S<O | .58 | .97 | .97 | .57 | .79 | .95 |

On S>O and S=O sentences, the English-speaking children at age 2;6 had SO interpretations more often than the German children (86% vs. 71% for S>O, and 78% vs. 67% for S=O). No difference between the English- and German-speaking children was found at older ages for either of these sentence types. Regarding the S<O sentences, the English-speaking children at 2;6 had SO interpretations just as often as the German children, both just under 60% of the time. At age 3;6, the English-speaking children had SO interpretations more often than the German children (97% vs. 79%). No such difference was found between the children at 4;6 for the S<O sentence type. Thus, by comparing how well word order was used for interpreting each sentence type at each age, we see that the English-children are generally ahead of the German-speaking children.

Children of both languages showed the same S>O >> S=O >> S<O pattern of performance at age 2;6—a pattern that disappears by age 3;6 in English-speaking children, but lingers in the German children of the same age. The observation that German-speaking children do not come to rely on word order until a later age than English-speaking children can be explained by a difference in properties of their target language. If it is assumed that a sufficient amount of positive evidence is necessary to result in the correct ranking of constraints (Tesar & Smolensky, 1998:238), then the speed at which children arrive at the correct ranking will differ if the input differs. For children to learn that word order is more important than animacy, they must—in a context in which the intended meaning is entirely clear—hear sentences in which SO order is used with inanimate subjects and animate objects. Once a child has encountered a sufficient number of these types of sentences in unambiguous situations, their grammar will make adjustments so that the animacy constraint is no longer incorrectly too high in the hierarchy. Children acquiring German encounter SO order less frequently than children acquiring English, so they are slower to reach the correct conclusion about the status of word order over animacy in their target language. Since German and Dutch both provide less positive evidence than English, the influence of animacy is predicted to disappear sooner in the English-speaking children than in Dutch-speaking children.

In this section we saw that the pattern exhibited by English-speaking children can be extended to other languages, German in particular. Because the input to the Dutch children patterns with the input to the German, we can predict that Dutch children at age 2;6 will show the same pattern that English-speaking and German-speaking children do at that age: S>O >> S=O >> S<O. Furthermore, we can expect that this pattern will persist in Dutch children even at the age of 3;6, as it did in German children of the same age in Chan et al. (2009). I also made a note of the fact that the sentences the children will be encountered with are not naturalistic, but are necessary to test their knowledge of word order in the face of animacy conflicts.

## 2.5 Summary

The goal of the present research is to answer the questions: is there an asymmetry in early subject-object word order, and if so is comprehension systematically affected by the relative animacy of subject and object? In the first section of the chapter, studies were reviewed that found variable comprehension in English-speaking preschoolers, yet other studies show that production of word order by young children appears to be unaffected by animacy contrasts. The data of one study in particular by Chapman and Miller (1975) stands out because the same set of preschoolers were tested both on comprehension and production of word order, and an asymmetry was found. However, more conclusive evidence is needed since the majority of studies since then have confounded animacy information with verb selectional restrictions. The present research aims to provide conclusive evidence by testing the same group of children in both comprehension and production, using sentences that are fully reversible.

A framework was established in the second section of the chapter as a tool for modeling a word order asymmetry. Frameworks unable to account for both variable comprehension and adult-like production were ruled out in favor of Optimality Theory, which treats comprehension and production as two different "directions" of the same grammar. The nature of the hierarchical system of soft constraints, some of which promote symmetry and others which promote asymmetry, allows for a comprehensive account of the early word order asymmetry. Hendriks, de Hoop, and Lamers' (2005) proposal involves an animacy constraint that is incorrectly ranked above a word order constraint. Because the animacy constraint plays no role in production, asymmetry is modeled. The model predicts that S>O sentences are easier to interpret than S<O sentences, while S=O sentences likely fall somewhere in the middle due to the fact that children are in the process of gradually reranking their constraints. Following de Hoop and Lamers (2006), a model was created to predict what type of processing difficulties S=O and S≠O sentences may incur in real time based on qualitative differences in constraint violation patterns.

The third section of the chapter addressed why word order, and more importantly, animacy can be treated as a universal constraint. In a framework that assumes universality of constraints, the fact that there are languages in which a violation of PROMINENCE results in ungrammaticality motivates the existence of a such a universal constraint. The animacy constraint is, therefore, proposed to exist in the grammars of speakers of languages like English, Dutch, and German, albeit with a low ranking relative to word order and other types of information used when interpreting transitive sentences. The animacy constraint is not expected to play a role in production, since word order alternations are largely the result of discourse prominence rather than inherent prominence (animacy).

The fourth section of the chapter looked at how the predictions of the Hendriks et al. model (2005), intended to model the acquisition of both English and Dutch, might be realized at different age groups for the different languages. Based on results from a study by Chan, Lieven, and Tomasello (2009) that tested English- and German-speaking children, I predicted that early word order in Dutch should be expected to pattern with early word order in German. Namely, Dutch children are expected to exhibit the pattern of performance S>O >> S=O >> S<O at age 2;6 like English-speaking children, a pattern that should also persist for the Dutch children until age 3;6 since word order is not as reliable a cue in their target language compared to English. Before jumping into the experiments to see if these predictions are met, I first address in the next chapter details about the experimental methods, design, and materials applied in chapters 4 – 8.

# 3 Testing for asymmetry

## Methods of assessing production and comprehension

The assumption was made in the introduction that psycholinguistic experimentation tells us something about children's and adults' linguistic knowledge. The purpose of this chapter is to describe and justify the experimental methods and design used in Chapters 4 – 7. It begins with an overview of the advantages and disadvantages of several methods for assessing very young children's language production and comprehension. For each method, a description of how it is used in the current paradigm is summarized. Next, several crucial controls of the experimental design and materials are discussed. The chapter ends with a summary alongside an outline of general predictions: how are preschoolers and adults expected to perform in light of the methods described in this chapter and the OT models discussed in the previous chapter.

Why give so much attention to the methods used if they are common methods in assessing children's linguistic abilities? First, if we are seeking to confirm or reject the hypothesis that there is an asymmetry of S-O word order, an explicit description of the methods used is merited. If children perform well on production, but poorly on comprehension, it must be clear that this is not due to the particular type of comprehension task used. Therefore it is important to justify the use of certain comprehension tasks with very young children and, moreover, to employ several different types of comprehension tasks. This may then rule out the possibility that children's performance is determined by the task rather than their knowledge. In other words, if the same pattern of variable performance is found across different comprehension tasks, it is fair to claim that children have incomplete knowledge or competence.

Second, I wish to show that for each method used, care was taken in the development of the experimental design. It was argued in Section 2.2.1.1 that an experimental limitation explanation may be too quickly appealed to to account for poor or variable performance on comprehension or production tasks; however, alterations of a task and design have been shown to improve the performance of children. An excellent example is the truth-value judgment task, used with children of at least age four, which requires participants to judge whether a sentence matches a picture or situation. Children give more adult-like responses if the experiment satisfies the condition of plausible dissent—that is, if the action corresponding to the false interpretation of a sentence was at some point a potential outcome in a picture or story (Conroy et al., 2009; Crain & Thornton, 1998; Crain et al., 1996). Children also appear to benefit from the opportunity to answer in a graded fashion via a magnitude estimation scale (Davies & Katsos, 2010), from being trained to detect infelicitous statements (Papafragou & Musolino, 2003), and from felicitous experimental contexts (Gualmini, 2004; Papafragou & Musolino, 2003). For these types of paradigms a distinction must also be made between true and false responses in matching and mismatching circumstances, since children are likely to answer *true* or *yes* to adults in general, especially if they are in doubt (Grimshaw & Rosen, 1990). What this example of the truth-value judgment task demonstrates is how important it is to tailor the experimental design to the particular task used. In what follows, I assert that my experimental design is appropriate for the tasks used with preschoolers in the present studies, namely elicited production, act-out, preferential looking, and picture selection with eye tracking.

## 3.1 Testing production

Two primary methods of investigating language production of very young children are discussed in this section: spontaneous speech collection and elicited production tasks. Because spontaneous speech is not particularly suited for the present investigation, an elicited production task is used alongside each of the different comprehension tasks in the present study. I give a general description of what the elicited production tasks entail. I then introduce the method of collecting gaze data via automatic eye tracking before discussing its application during sentence production.

### 3.1.1 Spontaneous speech

*Spontaneous speech* collection is a method of observation rather than actual testing. Speech data of children (and their adult or sibling interlocutors) is recorded and transcribed in natural settings, often in the home of the child or in a lab playroom. Available on the Child Language Data Exchange System (CHILDES) are transcribed speech data from children of various ages and languages (MacWhinney, 2000). Among the many advantages of this approach is that it can be used to investigate a variety of phenomena. It also provides information about frequency, both in child and child-directed speech. The method is particularly handy when investigating very young children since they are not met with any extraneous demands and are able to talk freely; there is also no real minimum age at which speech collection can begin since even babbling and partial words can be phonologically transcribed. The downside of using transcribed speech data is that it is usually available for only a small number of children, the researcher has little control over what is said, and the presence or absence of a construction cannot always be considered straightforward proof of the productive presence or definitive absence the construction (Demuth, 1996; Eisenbeiss, 2010; Stromswold, 1996).

While it is advisable to use spontaneous speech data in combination with experimental data to paint a complete picture of children's competencies, the present investigation requires an experimental setting to determine how preschoolers use S-O word order in their own speech. Transcripts are often unaccompanied by video recordings, which makes it difficult to determine the exact event the child is describing. This information—crucial in identifying word order as correct or reversed—is not always clear from a transcript alone. Experimental settings not only allow researchers to control the situations and events (i.e. intended meanings) they would like children to describe, but they also facilitate the direct comparison of production and comprehension.

### 3.1.2 Sentence elicitation

There are several ways to elicit language from experimental participants. Participants may be asked to complete a sentence that is begun by an experimenter or puppet. Some paradigms call for the repetition of a sentence, in which a sentence spoken by an experimenter or puppet must be imitated. *Sentence elicitation* is a sort of intermediary between elicited imitation and spontaneous speech collection in that structures of interest are elicited but not modeled directly (Eisenbeiss, 2010; Thornton, 1996). For example, a child is told a story by a puppet and then instructed to ask him a question or to answer the puppet's question. Simple transitive sentences with a subject and object are elicited from participants in the current study by means of an action or animation description task: the participant sees an action carried out by an experimenter or an action depicted in a cartoon animation, and is asked to describe it.

The major advantage of this method is that the meaning in the experimental context is controlled: the experimenter knows the exact event that is being described. Also, a robust data sample can be collected in a single session—as opposed to corpus data, which may not have a sufficient number of occurrences of a form for sound conclusions (Thornton, 1996: 78-79). However, a controlled setting does not guarantee that a participant will produce a target-like construction, or even the expected non-adult-like constructions. Think of the child who discusses the fact that the car is riding on the grass, what he finds to be a noteworthy breaking of conventional rules, rather than simply saying either "The car is pushing the cow" or "The cow is pushing the car." Another disadvantage is that there is a minimum age at which elicited production can be used. Children at age one or two cannot be reliably tested using controlled production experiments (Eisenbeiss, 2010: 22), and children at the age of two-and-a-half can be tested only "with effort" (Thornton, 1996). Indeed, we will see that young children often produce sentences in which word order is unable to be scored for various reasons, or they may refuse to respond to the task entirely.

The elicited production tasks used with preschoolers in the current study were designed so that they were engaging and so that felicity conditions were met. Because the children were not allowed to touch, move, or point at anything during the production task, in contrast to the more interactive comprehension tasks, it can be considered "less fun" than the comprehension task. McClellan, Yewchuk, and Holdgrafer (1986: 113) reported that several children "showed a marked preference for the comprehension task, as evidenced by their requests that the production task be terminated or that they be allowed to demonstrate the actions themselves." Thus, a friendly puppet animal was commissioned to interact with each child during the production task to give positive feedback and motivation in the form of praise, high fives, and snuggles. In addition, the role of the puppet in the game results in a felicitous experimental context—particularly important for children under the age of six (Thornton, 1996: 86). Since the assistant and child can both see the events, it would be infelicitous to ask the child to tell the assistant what is happening. For this reason the puppet closes its eyes and turns away, making it the child's task to tell the puppet what is occurring in the performed action or animation.

The elicitation task measures the proportion of SO and OS word order used in children's produced sentences. It is, therefore, essential to elicit sentences in which there are two NPs and a target verb. To increase the chance of eliciting such scorable responses from children, the production task always follows the corresponding comprehension task. A child who first hears over ten transitive sentence frames with two full NPs in a comprehension task is likely to be primed to produce the same type of sentences during a production task. Importantly, the sentences are presented during the comprehension task in the context of two or more possible interpretations. In this way, the meaning indicated by word order is not taught. Rather, sentences in which word order can be scored are encouraged.

*Rationale behind eye tracking*

Some of the production tasks are integrated with a technique of collecting data online, that is, as the produced sentence unfolds. In addition to the offline measure of word order used to describe an animation, participants' eye movements are measured as they view the animation on a remote automatic eye tracking monitor. As opposed to more "technologically stripped down" versions of eye tracking, in which eye gaze is monitored by a video camera and later coded by hand, automatic eye tracking uses a sophisticated system to register eye movements in real time (Sedivy, 2010: 128). An initial calibration procedure is then required so the system can relate reflections from the eye to a particular location on the screen; this is done by having the participant follow with their gaze a moving target that stops at various areas on the screen.

A remote system can be used, in which the eye camera is embedded in a computer monitor since very young children are not inclined to wear a head-mounted apparatus.

Before talking specifically about the manner in which gaze data can be collected during production tasks, it is useful to address the reasons it is believed that eye tracking is a valid psycholinguistic method. Sedivy outlines five assumptions underlying the use of eye tracking to the study spoken language, which are relevant for both production and comprehension (Sedivy, 2010: 116-118).

- First, people direct their gaze towards things they are attending to in their visual environment. As we "take in" a scene, several discrete eye movements called saccades (20 ms – 60 ms) are being made per second, of which we are generally not consciously aware. Each saccade is followed by a fixation, during which the eye is kept still for at least 150 ms (Trueswell, 2008). Saccades occur based on external factors of the stimulus like luminance, salience, or movement, as well as on internal, cognitive factors that determine what should be given attention.

- Second, eye movements are linguistically mediated. Sedivy offers examples from sentence comprehension, in which auditory linguistic input has been shown to be one of the internal cognitive factors that drives saccadic behavior, i.e. people's eye movements are driven by the goal of establishing reference (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). The same assumption appears to hold for production as well. Experiments in which participants are asked to describe a scene have shown a time-locking between eye movements and the ordering of constituents in utterances (Griffin & Bock, 2000; Papafragou, Hulbert, & Trueswell, 2008).

- Third, it is assumed that gaze reflects the incremental aspect of language processing. From the point of view of comprehension, a measurement like reaction time, which corresponds to the length of time it takes for decisions to be made about which participants are fairly certain, is temporally different from "cheap" eye movements, which correspond to "partial commitments" to interpretations as sentences are heard (Sedivy, 2010: 117). For instance saccades are made toward objects in the scene that are potential referents for a word within a few hundred milliseconds of the onset of the word (Allopenna, Magnuson, & Tanenhaus, 1998), and even anticipatory looks are made to objects that are likely arguments for a verb encountered (Altmann & Kamide, 1999). Likewise with production, eye movements reflect incremental processing. In a study by Gleitman, January, Nappa, and Trueswell (2007), it was shown that people will begin their sentence with the first thing that they (are forced to) attend to in a visual scene.

- Fourth, Sedivy stresses the fact that in sentence comprehension the link between language and eye movements are *not* shallow associations between words and objects in the visual scene. Rather, eye movements reflect how people are interpreting syntactic structure, as was demonstrated in a study by Spivey, Tanenhaus, Eberhard, and Sedivy (2002). When presented with a temporarily ambiguous sentence like *Put the apple on the towel into the box*, participants momentarily cannot know if *on the towel* is a locative prepositional phrase (where they are to put the apple) or a modifying prepositional phrase (which apple is being referred to). In a display with a single apple on a napkin, participants look to an empty towel as a possible location for the apple before ultimately placing the apple into the box. In a display with two apples—only one of which is on a towel—participants do not look to the empty towel as a possible location since the context increases the likelihood that the prepositional phrase is a modifier. In other words, just because one hears *towel*, one does not start looking at all towels in the visual scene; rather, there are deeper, linguistic motivations for gaze behavior.

- Fifth, if eye tracking is used with children, it must be assumed that the previous four assumptions also hold across a child's development. According to Sedivy, there is reason to believe that children's gaze behavior is attention bound, referentially driven, incremental, and sensitive to linguistic structure (Fernald, Zangl, Portillo, & Marchman, 2008; McMurray & Aslin, 2004; Nadig & Sedivy, 2002; Snedeker & Trueswell, 2004; Trueswell et al., 1999). However, some differences between children and adults must be kept in mind. Although basic ocular abilities are in place by the time children are a year old (Colombo, 2001), children exhibit a greater latency than adults in the launching of saccades. While it takes about 250 ms for adults and children as young as 12 years of age to launch a saccade towards a visual target, it takes about 450 ms for children at about 4-and-a-half years old (Yang, Bucci, & Kapoula, 2002). Also, children under the age of three are more susceptible to the distraction of external visual factors like motion or sudden onsets than older children and adults (Scerif et al., 2005). According to Trueswell (2008) if care is taken in experimental design to control for these factors known to capture attention, inferences about even young children's language based on eye tracking measurements should be possible.

Thus, the use of gaze data requires a number of assumptions about the link between eye movements and cognition, as well as about the continuity between adult and child ocular and linguistic systems. The method is overall very attractive since it allows for a comparison of off- and online response patterns, and it reveals the types of information children use as they process language. The major downsides to the method are practical in that the automated equipment is costly in terms of money and the manual method is costly in terms of human labor. Data loss is also a matter of concern when testing active children (Sedivy, 2010).

*Eye tracking during sentence elicitation*

Returning to the particular case of collecting gaze data during elicited sentence production, the method is able to tell us something about the online production of sentences. Adult speakers of English have been found to look first to the agent and then to the patient when asked to describe a picture depicting a transitive action (Griffin & Bock, 2000). In paradigms in which the speaker's attention is initially drawn to one of two participants of an action (e.g. via visual saliency or a subliminal attention-capture technique), the speaker usually begins the sentence with the initially attended to participant—even if this ultimately requires a passive sentence structure (Gleitman et al., 2007; Griffin & Bock, 2000). These findings suggest that people first attend to what they intend to be the subject (i.e. the agent), unless external factors cause them to attend first to something else, in which case they go ahead and begin their sentence with that and deal with it in whatever way is necessary syntactically.

Further evidence that speakers first attend to what they need to properly create a sentence comes from an experiment that investigated cross-linguistic differences in motion action description. English-speakers usually use verbs describing manner ("He skates to the snowman") and Greek-speakers usually use verbs describing path ("He goes to the snowman (on skates)." Papafragou, Hulbert, and Trueswell (2008) found that in an elicitation task, English-speakers first looked to the manner and Greek-speakers first looked to the path of the depicted action. Interestingly, in both this study and that of Bock and Griffin (2000), when the linguistic task was removed, eye movements showed no particular pattern; participants merely scanned the scene (Griffin & Bock, 2000; Papafragou et al., 2008).

Eye tracking data comes in the form of an (x,y) coordinate on the screen, corresponding to an eye fixation location, which is recorded several times per second. For each item, Areas of Interest, or AOIs are drawn over the different objects on the screen and defined, for instance, as either *agent* or *patient* of the action. A critical moment usually measured in production studies is voice onset latency, or VOL, which is the time between the presentation of the visual stimulus and the beginning of the speaker's sentence (Bock, Irwin, & Davidson, 2004: 263). The VOL for each item can then be used to synchronize the gaze data to each elicited sentence's onset as well as give an indication of how much planning (in milliseconds) is necessary before beginning to speak. It is helpful to use gaze plots to show the proportion of looks to the different AOIs over time, aggregated over participants, for each of the experimental conditions.

To sum up this section about testing production: when investigating language production in preschoolers, the methods of spontaneous speech collection and elicited production are available, but only the latter is suitable for testing S-O word order in a controlled way. It has the advantage of allowing for the collection of many examples of a linguistic construction, but always bears the risk of providing many unscorable responses, especially with young children. The elicitation task allows responses to be collected in the form of offline word order used in sentence. Participants are additionally eye tracked to see which entities in the visual scene they attended to with their gaze prior to speaking and as their own sentences unfold. The use of gaze data is merited since the necessary assumptions have been made: that both adult's and children's visual attention is related to linguistic processes.

## 3.2 Testing comprehension

In comparison to production, there is greater variety of comprehension tasks available for assessing the word order of children aged two and three. Although children under age four cannot reliably perform the popular truth-value judgment task or meta-linguistic grammaticality judgment task, the remaining options are not limited. This section reviews tasks that are able to be reliably performed by children aged two-and-a-half and three-and-a-half, namely the act-out, picture selection, and preferential looking tasks. The collection of gaze data via automatic eye tracking is additionally discussed in the context of preferential looking tasks.

### 3.2.1 Act-out task

In an *act-out task*, a participant hears a spoken or pre-recorded sentence and is asked to act out his or her interpretation of the sentence using a set of props or toys. Several examples of this type of paradigm were given in Chapter 2, including C&M's study in which children heard a sentence like *The boy is hitting the car* and had to act it out with two toys. A variant of the act-out task is an enactment task, which involves the child as a participant of the action. In a study investigating the acquisition of object scrambling in Dutch, for example, children were given items and asked to perform an action on them, like "Roll a marble twice" (Krämer, 2000). In the two experiments that will be reported in Chapter 4, the traditional act-out task was utilized: children were presented with two toys and asked to use them to carry out the transitive action they heard described.

It is favorable to use act-out tasks for several reasons. The game allows children to give their interpretations of sentences without a pre-determined range of interpretations. It is also seen as particularly fun and engaging, at least by children who are not too shy (Goodluck, 1996; Schmitt & Miller, 2010). While three-year-old children are claimed to be the youngest to be reliably tested with this task, several studies (like those reviewed in Section 2.1.1) have made conclusions about the language of two-year-olds using the act-out task (e.g. Bever, 1970; Chan, Meints, Lieven, & Tomasello, 2010; Chapman & Miller, 1975; McClellan, Yewchuk, & Holdgrafer, 1986; Thal & Flores, 2001). Furthermore, it is particularly useful in an investigation of agent-patient relations since it gives a clear indication of who is doing what to whom. This is especially the case when investigating such relations in actional transitive events, which are easy to demonstrate using props, as opposed to abstract mental or emotion verbs (Goodluck, 1996).

The downside of using an act-out task, especially with children aged two and three, is that it can be too free. Since the child is not limited to a set of possible responses, actions may be difficult to score or even completely non-target-like. And despite being a fun task, it is nonetheless cognitively complex in that it requires working memory, sentence decoding, and action planning (Goodluck, 1996). Because of this, children who find the task difficult may exhibit biases like performing actions on the props themselves or exhibiting side biases (Shatz, 1978). The task must also be carefully administered since children have also been shown to perform differently depending on whether they begin their response before or after the end of the auditory sentence (Goodluck, 1996; Meroni & Crain, 2002; Trueswell et al., 1999).

The act-out task used with Dutch preschoolers presented in Chapter 4 is designed to maximize success. Not success with the interpretation of word order *per se*, but with giving responses from which their interpretation of word order could be scored. In order to reduce the difficulty of the task and to limit unclear responses, the child is shown precisely how the toys can be used to act out the two test verbs (*push* and *pull*). Word order is not modeled, but rather the verb is named ("This is 'pushing'") and the action is modeled with one set of toys by the assistant. The child is then encouraged to carry out the action with a second set of toys. To increase the reliability of the responses to test sentences, the child is not given access to the toys until after a sentence has been presented once. The toys are physically kept out of the child's reach on a tray until the sentence is complete. In addition, only the two necessary toys are provided per item, rather than an array of toys from which two must be chosen. Regarding felicity conditions, the "do-what-you-hear" context is a natural game for children, usually eager to manipulate the toys. No gaze data is collected during the act-out task.

The design is meant to get the most out of the children, not to get *even more* out of the children. C&M, whose participants were as young as 1;8, re-tested items in a second and sometimes third session for which an unscorable response had initially been given. This seems at first like a desired luxury, to have a second chance at getting a scorable response, but it is actually not optimal to give some children two chances on items and others only one. On the same note, Bever and his colleagues, among the first to administer an act-out task, engaged in a extension of the task whenever children failed to respond (Bever, 1970; Bever et al., 1970). In the case that a child would not answer "the experimenter presented the child with two alternatives (acting them out herself) and asked the child which was the one that the sentence described" (Bever, Mehler, & Valian, 1970: 5). Without getting into obvious flaws (e.g. the order of presentation of the two alternatives were likely not reliably counterbalanced), it should be pointed out that this type of adaptation marks a reluctance on the part of the experimenter to accept a non-response for what it is. As frustrating as unscorable and non-responses are, attempts to extract more than this often means risking a biased result.

The act-out task measures the proportion of SO and OS interpretations children have, reflected by their actions. Only final responses in which the target or reversed action was performed (with both agent *and* patient) are counted as scorable answers. This is in contrast to Thal and Flores' (2001) scoring procedure:

> [ . . . ] agent selection was based on the first item the child moved, picked up, or touched regardless of the action performed and or the inclusion of a patient in the event sequence. A score of 1.0 was given if the child first moved, picked up, or touched the first noun mentioned. A score of 0 was given if the child first moved, picked up, or touched, the second noun mentioned. A score of 0.5 was given if the child moved, picked up, or touched both objects simultaneously or if s/he did not respond to the stimulus at all. (Thal & Flores, 2001: 181)

In Thal and Flores' scoring procedure, the child does not have to actually perform the action of, for example, hitting or kissing in order to get full credit for interpreting the meaning of word order. However, the simple touching or moving of a single toy does not indicate enough about how that child has interpreted a sentence. After all, it is not unlikely that a child might first reposition the toys before carrying out the target action. It also seems odd to give a child a score of 0.5 for simply doing nothing. A non-response should more conservatively be treated as missing data. Results based on a more conservative scoring procedure can be seen as highly reliable since they are based on responses in which the child has clearly expressed his or her interpretation of the sentence.

### 3.2.2 Picture selection

In a *picture selection* task, also referred to as a picture matching task, participants are presented with a linguistic stimulus in the form of either a word or a sentence and are asked to select the matching picture from a set. This method can been used to assess comprehension of phonological distinctions, lexical comprehension, and the effects of morphosyntactic manipulations on sentence comprehension (Gerken & Shady, 1996). In the experiments reported in Chapter 6, participants are presented with two animations on a screen and asked to select the one that correspond to the transitive sentence they hear. The target animation corresponds to a SO interpretation, and the distractor picture corresponds to an OS interpretation.

The picture selection task is different from the act-out task in that the set of interpretations of the sentence is limited in picture selection. The picture selection task has an advantage of being a straightforward and easy task. It has been successfully used with children as young as age two (Schmitt & Miller, 2010: 44). However, care must be taken that the saliency of each of the pictures is controlled, or else participants may choose one picture over the other on the sole basis of how bright, colorful, or interesting it is compared to the alternative (Gerken & Shady, 1996; Schmitt & Miller, 2010). Materials must also be checked to ensure that the event portrayed is clear, since it is not easy to clearly depict motion in pictures and photos. Very young children also have difficulty interpreting the meaning of curved lines around joints, a conventional way of indicating motion in illustrations (Cocking & McHale, 1977; Friedman & Stevenson, 1975). Another issue for concern is that children may give unusual responses, despite the simplicity of the task. They may fail to point, point at one picture then the other, or even point at both pictures with two hands.

In the picture selection task used in the present study, children are presented with a sentence and asked to point to one of two animations, presented side-by-side on a computer screen. Cartoon animations are used in order to effectively depict the transitive events of *pushing* and *pulling* and to simultaneously aid in keeping the children's attention. Practice items in which word order is not contrasted are used to introduce the task, for example, a bear on a bike is presented next to a cat with a bouncing ball, and the child must point to the animation matching *The ball is bouncing over the cat*. Difficulty increases over the three practice items, to make sure the child is paying attention to both the nouns (*The balloon is carrying the ball* vs. *The balloon is carrying the bottle*) and the verb (*The monkey is tickling the bear* vs. *The money is pointing at the bear*). The task measures the proportion of SO and OS interpretations children have, reflected by their animation selections during test items.

### 3.2.3 Preferential looking

In a preferential looking task, in contrast to the act-out and picture selection tasks, no conscious response is required of the child. As with the picture selection task, the linguistic stimulus in a preferential looking task matches only one of two visual stimuli; rather than pointing, the child is simply expected to look longer at the matching stimulus if he or she comprehends the linguistic stimulus. The procedure has been used in investigations of lexical development, S-O word order, pronouns, prosody, and the meaning implications of verb frames (e.g. Hirsh-Pasek & Golinkoff, 1996a; Naigles, 1990).

The advantage of preferential looking is that it can be used with children as young as 12 months. In fact, evidence for comprehension is generally found using the paradigm at earlier ages than most other assessments (Hirsh-Pasek & Golinkoff, 1996b: 120). However it is often the case that many participants must be tested in order to obtain enough cooperative, attentive children—a fact that carries with it the possibility that the unengaged children do not understand the linguistic stimuli (Hirsh-Pasek & Golinkoff, 1996b). In order to be suitable for young children, a trial in a preferential looking study typically begins with an attention getting female voice that remarks “Wow!” or “What’s going on?” as the child is presented with the visual stimuli. The stimulus word or sentence can be repeated so as to prevent trials from being too swift and disorienting.

*Eye tracking during preferential looking*

As advanced technology has become increasingly available over the years, the *preferential looking* procedure has been adapted accordingly. The task was first formalized by Hirsh-Pasek and Golinkoff (1996b; 1996a), who dubbed it the intermodal preferential looking paradigm (IPLP). The classic IPLP involves two separate monitors located about 30 inches apart, before which an infant (and blindfolded caretaker) are seated. A button box is used for coding the infant’s looks, which entails a 300 ms latency—the time it takes for the observer to press the coding button. Fernald and colleagues later developed the so-called looking-while-listening paradigm, which increased the precision of coding (Fernald et al., 2008). The looking-while-listening procedure introduced a higher precision manual coding process. Eye movements in this paradigm are coded from videotapes in slow motion, supplying a data point every 33 ms. This improvement is essentially the same as “poor man’s eye tracking,” the technically stripped down and labor intensive eye tracking method mentioned in Section 3.1.2. Now that child-friendly remote systems are available, preferential looking is possible using automatic eye tracking technology (Johnson & Zamuner, 2010: 85).

As a result of the technological improvements, the line between preferential looking and regular eye tracking procedures used with older participants has been blurred. Preferential looking procedures were developed for very young children and usually involve two competing visual stimuli, while regular eye tracking procedures used in comprehension studies with older children and adults typically involve a single visual scene with multiple possible referents. A distinction is often made between the two on the grounds that preferential looking typically involves global measures of fixation time towards a target, rather than the time course of gaze to critical moments in the linguistic input. However, if an “eye tracking technique” can be defined as “the use of continuous eye monitoring during speech in such a way that allows for a fine-grained temporal analysis of eye gaze grounded with respect to specific points in the speech stream” (Sedivy, 2010: 115-116), then a preferential looking task can now fall under this category. The increased sensitivity of the measurements recently introduced to the preferential looking paradigm by manual and automatic eye tracking allows for a fine-grained time course and has consequently decreased the size of traditionally large analysis windows (Fernald et al., 2008). The relevant differences that remain between preferential looking and eye tracking are that preferential looking necessarily lacks an accompanying offline task, and the young target group requires an attention-maintaining trial design.

The preferential looking task presented in Chapter 6 involves the exact same stimuli and trial design used in the picture selection study. Rather than being asked to select one of the animations by pointing, the children are instructed to simply watch. Because children under the age of three are susceptible to being guided by visual attractiveness rather than by linguistic calculations, tracking gaze during a baseline without any linguistic stimulus is a useful way to control for potential visual confounds (Scerif et al., 2005; Sedivy, 2010; Trueswell, 2008). Thus, a 2.5 second baseline of visual input is presented, followed by 7 seconds of the same visual input accompanied by linguistic audio input. In order to maintain attention, trials begin with a fun sound together with an attention getting video (of a light bulb or rattle) at the center of the screen. Whether it is a baby laughing, a whistle, or Elmo's voice, children cannot help themselves from returning their attention to the screen upon hearing the sound. To keep a comfortable pace, the experimental sentences are repeated, with an excited interjection in between (*Wow!*, *Look at that!*, etc.). Though intended to make the trials appropriate for the preferential looking task, these elements serve to keep the children's attention when presented in the picture selection task as well.

The traditional analysis carried out with preferential looking data is a simple comparison of total time spent looking at the target display to total time the distractor display, often over a region of over five seconds (Hirsh-Pasek & Golinkoff, 1996b: 114-115). Fernald and colleagues have defined new, more precise dependent measures made possible by the increased number of data points now able to be collected in the paradigm. The first is an RT measure assessed on only distractor-initial trials by calculating the latency of the infants' first shift away from the distractor toward the target picture, measured from a critical point in the sentence stimulus. The second is a sort of *looking accuracy*, or the mean time spent looking at the target display as a proportion of total time spent looking at either the target or distractor display, calculated over a particular region (Fernald et al., 2008: 124-129). In the present study, the RT measure of Fernald is not used since the design does not result in a distractor-initial trial. That is, the animation-pair and sentence are presented simultaneously, allowing no time for a pre-stimulus preference for one picture over the other.[7] Instead, looking accuracy is measured from the offset of the sentence's subject, in several regions of 1000 ms, to observe how the children's looking behavior during the task changes over the course of the test sentences. Thus, the preferential looking task measures the proportion of SO and OS interpretations children have, reflected by their looking behavior during test items.

*Preferential looking vs. other comprehension tasks*

As was mentioned at the beginning of the chapter, it is advisable to make conclusions about children's knowledge of a linguistic construction based on their performance on multiple tasks. It is not the case, however, that similar performance should necessarily be predicted on each task used. Results from studies that included the preferential looking approach with a second task have shown that children tend to exhibit greater comprehension on the preferential looking task. A study by Chan, Meints, Lieven, and Tomasello (2010) presented in Section 2.1.1 found that a single group of two-year-olds could understand word order in [+an +an] sentences with familiar verbs in a preferential looking task, but not in an act-out task. Likewise, in a study investigating German preschoolers' comprehension of verb number inflection, Brandt-Kobele and Höhle (2010) found that a group of two- and three-year-old children understood verb inflection in a preferential looking task, but a second group of three-year-old children could not do so in a picture selection task. (Nor did the second group of children have a preference for looking to the target picture *during* the picture selection task. See Sekerina, Stromswold, and Hestvik (2004) for an example of discrepancy between offline picture selection responses and online gaze during the picture selection in

[7] Any preference during the baseline would not have been appropriate to use for this purpose since there was a fixation cross in the center of the screen between baseline and trial.

older children.) Brandt-Kobele and Höhle (2010: 1923) point out the importance of defining what kind of performance is sufficient to determine that a child has competence:

> If one considers full comprehension capacity to require adult-like performance in all kinds of tasks that tap children's sentence comprehension skills, then we have to conclude that the 3-year-old children tested in our study did not show a full ability to interpret verbal inflection. On the contrary, if one argues that finding empirical evidence for children's correct interpretation in at least one of the methods available is a sufficient demonstration of the children's linguistic capacity, then variations across tasks have to be considered a consequence of specific processing constraints in children.

Chan and colleagues are inclined to have the former view, that children who do not show comprehension skills on all tasks have incomplete knowledge, or weak representations, following from the idea of representation strength of Munakata, McClelland, Johnson, and Siegler (1997). Brandt-Kobele and Höhle, on the other hand, are inclined to support the latter view, that children have knowledge but are prevented from exhibiting it due to the processing demands of a certain task, such as the necessity to make a conscious decision in the picture selection task.

In the case of the present experimental design, the preferential looking task can be seen as a control. Children are predicted to make mistakes on the act-out and picture selection tasks due to the S-O animacy manipulation. Results from a preferential looking task may determine whether these animacy effects result as an extra-grammatical strategy in the face of a difficult task or whether they are linguistic effects that will further be revealed during the decision-less looking task. If children show complete competence of S-O word order in the face of animacy conflicts during a preferential looking task, then according to Brandt-Kobele and Höhle's view a production-comprehension asymmetry is not supported (Brandt-Kobele & Höhle, 2010: 1922-1923).

To sum up this section about testing comprehension: when investigating language comprehension in preschoolers, the act-out task, the picture selection task, and the preferential looking tasks are available. Children's interpretations of test sentences are free in an act-out task with toys and their interpretations are bound to possibilities provided in a picture selection task. The preferential looking task in the present study is identical to the picture selection task in that two cartoon animations are presented side-by-side, but in the former the child is not required to point, but to simply watch the screen. Thanks to improvements in technology, the precision of automated eye tracking can be used to define a fine-grained time course of gaze during preferential looking to see how children's preferences change as a sentence unfolds. The present study aims to use all three tasks to draw as full of a picture of their grammatical competence as possible. Because the materials used across the tasks are controlled, results from the different tasks are able to be fairly compared, as discussed in the next section.

## 3.3 Experimental controls

The tasks described in the previous sections require somewhat different paradigms for testing word order of preschoolers. How can materials be controlled in order both to test the hypotheses about word order and to allow the results from the different tasks to be compared? In this section I first address important aspects of the experimental design in detail: the experimental sentences and the visual materials. Importantly, the experimental sentences were carefully controlled and were the same across all tasks. I then give reasons for why results can be compared across comprehension tasks, across comprehension and production tasks, and between children and adults.

### 3.3.1 Experimental sentences

The experimental sentences used in the present study appear in Appendix A: test sentences are listed in Table A.2 and practice and filler items are listed in Table A.4 – A.6. In the test sentences, four animals (or four humans) and four vehicles push and pull each other. List B is the reverse of List A: test item 1a is *de koe duwt de hond* (*the cow is pushing the dog*), and 1b is *de hond duwt de koe* (*the dog is pushing the cow*). In total, there are 16 sentences per list, each containing 4 items from each sentence type: [+an +an], [+an -an], [-an +an], and [-an -an]. An overview of the experiments using these sentences is given in Table 1.1 at the end of Chapter 1. It is only in Experiment 2 that humans are used instead of animals as animate entities.

Why were these sentences used? It is apparent from the review of studies following C&M in both English (Section 2.1.1) and other languages (Section 2.4.1) that the sentences used to investigate the influence of animacy on word order too often confound several types of information: S-O animacy, verb selection requirements, and event probability. In this section, it will be argued that the sentences used in the present study are sufficiently reversible to allow for a fair assessment of the pure effects of S-O animacy on word order. When creating the sentences, close attention was given to the verb argument selection properties and event biases, meanwhile keeping the limited vocabulary of preschoolers in mind. The level of animacy within the [+an] and [-an] noun sets was also controlled. Expectations about the default interpretation of these sentences is addressed at the end of this section, making reference to context, specificity, and prosody.

#### *3.3.1.1 Isolating animacy*

In the naturalistic speech that is directed towards children, animacy is a team player. Input from the target language includes sources of information that are correlated with word order and animacy. For example, many case-marked pronouns are used in child-directed speech to refer to humans, and many transitive verbs are used that are irreversible with respect to animacy, like *drink* or *read*. This means that animacy does not often work alone as a source of information.

Children encounter transitive sentences in naturalistic speech in which several of the five distinguishability constraints (case, agreement, verb selectional restrictions, word order, and animacy) are satisfied by a particular interpretation. This is demonstrated with a few samples from Dutch child-directed speech in the Groningen corpus from CHILDES (MacWhinney, 2000; Wijnen & Verrips, 1998). Sentences (1) – (3) have SO order, and SO interpretations of these sentences violate none of the five distinguishability constraints. Sentences (4) and (5) have OS order, and OS interpretations violate none of the five distinguishability constraints (except, of course, for the word order constraint). It is likely that children often hear sentences in which a certain interpretation is corroborated by multiple satisfied constraints.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| (1) | *Ik* | *heb* | *nog* | *geen suiker* | *erin* | *gedaan* | (SO) |
| | I.1.SNG.NOM | have.1.SNG | yet | [no sugar]NOM/ACC | in there | do.PST | |
| | 'I haven't put any sugar in there yet' | | | | | | |

(To Josse 2;04)

| | | | | | | |
|---|---|---|---|---|---|---|
| (2) | *Dan* | *maken* | *we* | *nou* | *de groene* | (SO) |
| | Then | make.1/2/3.PL | we.3.PL.NOM | now | the green one.NOM/ACC | |
| | 'Now let's do the green one' | | | | | |

(To Matthijs 2;03)

| | | | | | | |
|---|---|---|---|---|---|---|
| (3) | *Je* | *had* | *het koffieapparaat* | *aangezet* | | (SO) |
| | You.2.SNG.NOM/ACC | have.1/2/3.PST | [the coffee machine]NOM/ACC | turn.on.PRT | | |

‘You turned on the coffee machine’

(To Peter 2;02)

| | | | | | | |
|---|---|---|---|---|---|---|
| (4) | *Ja,* | *die* | *krijg* | *je* | *ook* | (OS) |
| | Yes, | that one.NOM/ACC | get.1/2.SNG | you.2.SNG.NOM/ACC | also | |

‘Yes, that one you get, too’

(To Abel 2;04)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| (5) | *Die* | *heb* | *je* | *zelf* | *gemaakt,* | *he?* | (OS) |
| | That one.NOM/ACC | have.1/2.SNG | you.2.SNG.NOM/ACC | by yourself | make.PST | hey? | |

‘That one you made by yourself, didn’t you?’

(To Tomas 2;03)

The present research aims to test the hypothesis that preschoolers rank animacy higher than word order by testing controlled experimental sentences. Unlike naturalistic speech, the experimental sentences will only provide word order and animacy information so that we can see which the children will use for interpretation.

*Case and agreement*

In the experimental sentences, the lack of case marking on the noun phrases (only possible on pronouns in Dutch and English) and the singular agreement marking on the verb (VERB-*t* in Dutch and *is VERB-ing* in English) result in sentences that are reversible with respect to AGREEMENT and CASE.

*Selectional restrictions of verbs*

In the experimental sentences, the verbs *push* and *pull* are neutral with respect to the kind of arguments they take. Otherwise it is not clear as to whether the animacy effect is due to the selectional requirements of the verb or to the relative animacy of subject and object. Remember that de Hoop and Lamers (2006) demonstrated using sentences from German that S-O animacy and verb selectional restrictions are two distinct types of information used to interpret word order (Section 2.3.2). That is, when SELECTION and PRECEDENCE are in competition in German, SELECTION wins regardless of whether PROMINENCE is satisfied or not; and when PROMINENCE and PRECEDENCE are in competition, PRECEDENCE wins.

Experimental sentences that do not tease apart verb selectional requirements and S-O animacy cannot tell us anything about the pure effect of S-O animacy. Sentences like (6) – (9) have been used in studies investigating the effect of “animacy,” but actually include confounding verb information.

| | | |
|---|---|---|
| (6) | The eraser licks the rabbit | (French: Kail, 1989: 86) |
| (7) | The cup is kissing the bear | (English: Thal & Flores, 2001: 179) |
| (8) | The boy opens the door | (English: Corrigan & Odya-Weis, 1985: 53) |
| (9) | The dog grabs the pencil | (English & Italian: Bates et al., 1984: 345; English, German, & Italian: MacWhinney, Bates, & Kliegl, 1984: 139) |

A child may interpret sentences (6) and (7) as OS, but one must be careful before concluding that it is based on the animacy of the NPs alone. Its more likely interpretation is based on the fact that *licking* and *kissing* require animate subjects rather than the fact that a rabbit or bear is more animate than a cup or eraser.

To stress the importance of this point, I share here the results of early studies of English-speaking children's and adults' sentence interpretations of active and passive sentences. An active sentence with a reversible verb like *follow* as in (10) is easier to interpret than a passive sentence with a reversible verb as in (11) (Bever, 1970; Gough, 1966; Slobin, 1966). The difference in difficulty between active and passive sentences decreases, however, when sentences are tested that contain verbs that are not reversible in light of their argument selection restrictions. In a truth-value judgment task in which RTs were measured, Slobin (1966: 225-226) found that neither children aged six to twelve nor adults found irreversible passives like (13) to be more difficult than irreversible actives like (12).

| | | |
|---|---|---|
| (10) | The turkey followed the pig | (reversible active) |
| (11) | The pig was followed by the turkey | (reversible passive) |
| (12) | The girl is watering the flowers | (irreversible active) |
| (13) | The flowers are being watered by the girl | (irreversible passive) |

Turner & Rommetveit (1967: 654) tested children aged five to eight on a similar task and found that children correctly responded to irreversible passives over 80% of the time by age five, but did not do the same for reversible passives until age eight. Thus, a sentence that is irreversible because of its verb selectional restrictions is easier to interpret than a sentence that is reversible with regard to those restrictions. Moreover, children are sensitive to the selectional restrictions of the verb, so irreversible sentences should not be used when investigating the pure effect of S-O animacy.

*Push* and *pull* were chosen for the present studies because they are verbs known by young children that take on both animate and inanimate subjects and objects. Table A.3 in Appendix A shows for each word in the experimental sentences the percentage of Dutch- and English-speaking children aged 30 months who can both understand and say it (Dale & Fenson, 1996; Zink & Lejaegere, 2002). Over 74% of children of both languages are familiar with the two verbs used in the test sentences. To ensure that children were aware of how the test verbs would appear in the experiment, the act of pushing from behind and pulling forward with a rope were demonstrated before each experimental session. Word order was not modeled during the verb demonstration (e.g. "This is 'pulling'"). Test sentences for *push* were presented in a separate block than test sentences for *pull* to avoid unnecessary confusion between the two.

#### *3.3.1.2 Event probability*

Closely related to verb argument selection properties is event probability. While sentences like (8) and (9) cannot be reversed due to the selection properties of *open* and *grab*, a sentence like (14) can be reversed because both the SO and OS meaning satisfy the selection properties of *chase*. However, it is still not ideal for testing the comprehension of S-O word order because one interpretation is more likely. That is, even though both interpretations of (14) are possible, one of the two is more likely to take place.

(14) The dog is chasing the cat

It can be the case that *cat* is interpreted as the object simply because dogs usually chase cats, and not because the word order is being paid attention to, *per se*. Thus, it is vital that there is not a bias for one or the other interpretation of a test sentence based on world knowledge. Otherwise the effects of linguistic information like S-O word order cannot be clearly assessed.

There is evidence that children are sensitive to event biases at an early age. Bever (1970) tested children aged two to five with an act-out task on their comprehension of simple active sentences like (15) – (17). He intended sentences like (15) to be fully reversible, (16) to be reversible but probable, and (17) to be reversible but improbable. Thus the preferred interpretation of (17) based on event biases would be the same as (16).

(15) The cow kisses the horse
(16) The mother pats the dog
(17) The dog pats the mother

He found that all of the children interpreted fully reversible and probable sentences correctly over 90% of the time. Improbable sentences posed greater difficulty. Children under age four interpreted improbable sentences correctly between 50% and 70% of the time; children over age four interpreted improbable sentences between 80% and 90% of the time. In fact, Bever (1970: 297-298) speculates that world knowledge usually overrides linguistic information in natural settings, even for adults. In non-experimental settings "specific contexts must provide [a] far stronger basis for prediction of the most likely meaning of a sentence independent of its form." Thus, he claims, it is likely that normal processing largely disregards actual ordering of words or syntactic structure. This idea is supported by Ferreira and colleagues who found "good enough" or "shallow" parsing of sentences by adults aided by event probability in the face of non-canonical sentence structure (Ferreira, 2003; Ferreira, Bailey, & Ferraro, 2002). As a consequence, sentences used to test the comprehension of S-O word order should be free of event biases that would encourage listeners to ignore linguistic cues like word order.

Event biases overrule not only word order, but also eventual effects of S-O animacy. Chapman and Kohn (1978) used an act-out task to investigate the effect of word order, animacy, toy position (agent location on left or right of patient), and event probability on preschooler's interpretation of sentences like (18) – (23). They defined *a priori* that (18) is more probable than (19) and that (21) is more probable than (20). That is, dogs tend to chase cars, and vehicles tend to hit people. They assumed that events in sentences like (22) and (23) were equally likely to occur.

(18) The dog chases the car
(19) The car chases the dog

(20) The boy hits the truck
(21) The truck hits the boy

(22) The girl bumps the swing
(23) The swing bumps the girl

Children aged three-and-a-half consistently relied on word order to act out the sentences. Two-year-olds, on the other hand, exhibited several strategies. Some had consistent position preferences (e.g. the toy on the right always became agent), which will be discussed further in the next section. Other children showed a preference for animate subjects and inanimate objects—but not in sentences in which the inanimate entity would be the probable subject. Thus S<O sentences like (23) would be interpreted by these children as OS, but not S<O sentences like (21) with an inanimate subject bias. Similarly, in an attempt to replicate C&M, McClellan, Yewchuk, and Holdgrafer (1986) found that production of word order preceded comprehension of preschoolers, but mistakes in comprehension revealed the reverse pattern than expected, with the best performance on the [-an +an] sentences. McClellan et al. largely used the same sentences as C&M, including those like (18) and (21).[8] Both Chapman and Kohn (1978: 759) and McClellan et al. (1986: 111) each concluded from their results that there is no general animacy effect since event probability can overrule it. However, if animacy information is assumed to be linguistic and event probability extra-linguistic, then the latter should not be used to disprove an animacy effect, but should instead be controlled for when testing for an animacy effect.[9]

Sentences tested in the present studies containing *push* and *pull* together with animals (or people) do not contain event biases. For example, cows do not usually push cars, nor do cars usually push cows. Animals who typically pull vehicles were not used (e.g. donkeys or horses). A study by de Swart and Cannizzaro (in preparation) suggests that there is no existing bias for animals to push or pull vehicles. In a materials pre-test, S≠O sentences were rated in which animals (e.g. *bear*, *lion*, *whale*) and vehicles (e.g. *tram*, *taxi*, *steamship*) were engaged in pushing and pulling each other. The S>O and S<O sentences earned equally low scores (around 2.3 out of 7) on a plausibility rating test completed by adults speakers of Dutch, despite prepositional phrases containing plausible settings for these unlikely events. Sentences like (24) were judged to be equally (im)plausible as sentences like (25).

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| (24) | *Vanochtend* | *heeft* | *de taxi* | *de leeuw* | *naar de circustent* | *geduwd* | (SO) |
| | this morning | has.SNG | the taxi | the lion | to the circus tent | push.PTC | |
| | 'This morning the taxi pushed the lion to the circus tent' | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| (25) | *Vanochtend* | *heeft* | *de leeuw* | *de taxi* | *weg van haar welpjes* | *geduwd* | (SO) |
| | this morning | has.SNG | the lion | the taxi | away from her puppies | push.PTC | |
| | 'This morning the lion pushed the taxi away from her puppies' | | | | | | |

Finally, and most importantly, two lists of test sentences were used (list B as the reverse of A). In this way, any biases that might exist are counteracted. For example, if an item has a bias in one direction (children expect boats to pull fish), than the same item in the reversed list should have a bias in the opposite direction (children do not expect fish to pull boats).

Event probability of specific events is related to, but not the same as, frequency of overall event types. While the former has to do with how likely a specific action is to occur between two entities, the latter has to do with general tendencies that may be the source of general linguistic preferences. When discussing the motivation of the animacy constraint (Section 2.3.1), we saw that sentences with S>O

---

8 The sentences in C&M and McClellan et al. (1986) differed from (18) – (23) in that they were in the present progressive. See Table A.1.

9 Bouma (2008: 194, 221-222) points out that there is no clear way of positing an event plausibility constraint in the OT grammar. If one were to exist, he suggests that it would hold in comprehension only and would have to be expressed in terms of predicate-argument frequencies. However, event plausibility calculated from events that adults and children hear expressed linguistically is inadequate, since it is likely also based on the frequency of events experienced that are not necessarily expressed linguistically. That is, a child knows that dogs usually chase cats or cars because the child has seen these events—events that may or may not have been additionally described. It will be assumed, therefore, that event plausibility is based on world knowledge and is not necessarily or sufficiently defined by linguistic frequencies.

meanings occur 86% of the time in Dutch adult speech (Bouma 2008: 257), and 95% in Dutch child speech (Hogeweg & de Hoop, 2010). While *push* and *pull* can take on animate or inanimate subjects and objects, the types of actions that Dutch children encounter with *push* and *pull* also usually involve an animate agent and an inanimate patient. Using the Groningen corpus from CHILDES containing 170 hours of speech from 7 Dutch children aged 1;05 – 3;07 and their parents (MacWhinney, 2000; Wijnen & Verrips, 1998), I inspected sentences containing a form of the verb *push* or *pull*. There were 120 sentences in child-directed speech (30% with push and 70% with pull) and 67 in child speech (40% with push and 60% with pull). For both parents and children, subjects were animate over 80% of the time and objects were inanimate over 85% of the time. Nevertheless, the infrequency of inanimate subjects used with *push* and *pull* and with verbs in general can be seen as the source for the preference for a certain type of meaning, not as an issue of the likelihood of a specific event. (Nor should the infrequency of inanimate subjects be seen as a source of an asymmetry since children appear to correctly order inanimate constituents before animate ones in their own productions.)

#### *3.3.1.3 Animacy categories*

Common animacy hierarchies, briefly sketched in Section 1.2.2, typically place humans higher than animals (Aissen, 2003; Comrie, 1989; Corrigan, 1986; Van Valin & Wilkins, 1996; Wilkins, 1990; Zaenen et al., 1976). For this reason, studies that involve [+an +an] sentences should not include humans and animals in the same [+an] category. Sentences (26) – (30) are examples of test sentences from studies investigating the effect of animacy on early sentence comprehension that treat humans and animals as equal on the animacy scale.

| | | |
|---|---|---|
| (26) | The dog is chasing the boy | (Chapman & Miller, 1975: 358) |
| (27) | The girl pushes the dog | (McClellan et al., 1986: 101) |
| (28) | The baby is being touched by the frog | (Lempert, 1989: 237) |
| (29) | The pig bumps the queen | (Lempert, 1990: 685) |
| (30) | The bear is hitting the girl | (Thal & Flores, 2001: 179) |

Childers and Tomasello (2001: 74) also treated humans and animals as equally animate entities. All animate entities in the current experiments were animals, with the exception of Experiment 2, in which humans took the place of the animals in the test sentences. Table A.3 in Appendix A shows that between 80% and 100% of both Dutch- and English-speaking children aged 30 months are familiar with the animate and inanimate nouns used in the experimental sentences. Each test session began with a naming pre-test to make sure the children were familiar with the names for the nouns used in the experiment.

Inanimate nouns in the test sentences are always vehicles (e.g. *car*, *boat*, *airplane*), while inanimate nouns in the practice and filler sentences vary (e.g. *ball*, *chair*, *moped*). Vehicles are chosen for the test sentences because they are inanimate but still satisfy the selection requirements of *push* and *pull*: they are capable of pushing and pulling other entities. While vehicles are not capable of independent motion, moving is their inherent purpose and the people who are operating vehicles are often not visible to an onlooker.

Chan, Lieven, and Tomasello (2009) tested the English- and German-speaking preschoolers on sentences with animals interacting with inanimate entities that are not as prone to movement as vehicles, like *telephone*, *present*, *house*. The novel verbs *tam* and *meek* that were used in the study each involved a special apparatus that allowed normally immobile inanimate entities to move another entity. For instance in the resting position of *tam*, both toys stood on the protruded part of a spherical toy that was able to move up and down. The *tammer* would push the *tammee* down, resulting in the *tammee* moving up and down (Chan, Lieven, & Tomasello, 2009: 279). Their design seems to allow for a greater difference in animacy between the [+an] and [-an] nouns than is possible with the verbs *push* and *pull*; however, one can never be entirely certain that children were exhibiting an animacy effect when they rejected a telephone as a *tammer* in favor of an animal, since it could also be the case that they simply did not find it feasible that a telephone would be capable of the motion entailed by *tamming*. Vehicles are, therefore, more adequate than immobile entities as [-an] agents in the present investigation.

#### *3.3.1.4 Context*

Linguistic context is often necessary for the proper interpretation of sentences. A pronoun, for example, requires a linguistic antecedent in the discourse. Furthermore, different types of contexts can strengthen or alter interpretations of sentences. Children are more likely to correctly understand a pronoun in felicitous (Conroy et al., 2009) or coherent linguistic contexts (Spenader et al., 2009). In the case of S-O word order in Dutch, certain contexts may result in an OS interpretation since OS word order is merited in contexts of topicalization or question answering. In the present study, the experimental sentences were delivered without any linguistic context.

The fact that the sentences were delivered with no linguistic context during the comprehension task is not problematic for the present investigation of S-O word order. First, context is not necessary for the interpretation of word order. Unlike sentences with pronouns, sentences with full NPs can convey meaning in isolation. Second, a sentence delivered out of context encourages the default interpretation of word order, which in the case of Dutch and English is SO (Kaan, 1997; see also Kaiser & Trueswell, 2004). In other words, topicalization does not occur since there is no discourse new or old information. Both NPs in a sentence that is presented are new (no noun occurring in an experimental sentence appears in a preceding or following item), and each sentence is pre-recorded with a neutral prosody. Thus, the optimal interpretation of each sentence is a canonical SO interpretation.

Participants hear sentences out of context, yet the two NPs contain definite articles, which may be seen as infelicitous. However, the sentences are always presented with a visual context. Because participants see a cow and a car in front of them as they hear *The car is pushing the cow*, it is not infelicitous to use a definite article. In addition, each of the entities that appear in the experimental sentences had been introduced to the children before the test session began, thereby increasing the felicity of using a definite article with an NP in this paradigm.

Since a lack of context points to a default SO interpretation, children's misinterpretation of word order should not be attributed to the lack of linguistic context. Even if there is a context that could be construed that would help children interpret S-O word order more often in an adult-like way, it does not affect the matter currently being investigated. After all, if a certain type of context improves children's performance on a comprehension task, it would not preclude an effect of animacy. A context effect would be seen as independent of any potential animacy effect since animacy is an inherent property of the subject and object, not the context. In short, adults do not require special discourse context to arrive at SO interpretations for simple reversible sentences, so no such context is provided for the children.

### 3.3.2 Visual materials

Similar measures were taken across all tasks to control for size saliency of the visual materials used. The verbs deemed acceptable earlier in this section, *push* and *pull*, can easily be acted out with toys and can be depicted in two-dimensional animations. The toys used in the comprehension and production tasks of Experiment 1 and 2 appear in Figures B.1 – B.6 in Appendix B; the animations used in the comprehension and production tasks of Experiments 3 – 6 appear in Figures B.7 – B.38. In order to keep all of the nouns equally prominent with regard to physical properties, the size of the entities in test items (whether as a toy or animation) is equal. In addition, each entity in the cartoon animations has one aspect that moves in a subtle fashion, for example the legs of the cow or the wheels of the car.[10] Keeping the movement subtle and equal between entities is important as a control for both eye tracking (Trueswell, 2008) and possible effects of visual salience on word order interpretation (Chapman & Kohn, 1978; Hendriks et al., 2005) or production (Griffin & Bock, 2000; Tomlin, 1995).

Measures are also taken across all tasks to control for side and direction biases. McClellan, Yewchuk, and Holdgrafer (1986) found that some two-year-olds in their act-out study investigating word order consistently assigned agent status to the toy located on a preferred side. It is likely that some children might also consistently point to (or look at) an animation that is on a preferred side of the screen. In the present act-out task, the location of the toy that represents the agent is counterbalanced across sentence types; in the picture selection and preferential looking tasks, the location of the target animation is counterbalanced across sentence types. Finally, whenever an action is depicted either for description in the elicited production tasks or on screen in the picture selection and preferential looking tasks, the direction of action (towards the left of the right) is also counterbalanced across sentence types. The counterbalancing of agent or target picture position ensures a 50% accuracy rate per sentence type for a child with a side bias. The counterbalancing of action direction in the production tasks prevents children from interpreting or producing sentences without fully paying attention to the action. For instance, if the direction of pulling is always from left to right, a child might learn to simply always choose the entity on the right as agent during comprehension or production instead of attending to who is pulling and who is being pulled.

### 3.3.3 Task pairing

Having established that the experimental sentences and materials are appropriate for testing early S-O word order, I now address how the experimental design allows for a fair comparison of (i) performance across comprehension tasks, (ii) performance on comprehension and production tasks, and (iii) performance by children and adults.

*Across comprehension tasks*

Children are tested on three comprehension tasks, of which two require a conscious decision: act-out and picture selection. In the act-out task, the child hears a sentence and has to assign meaning to it. With two toys placed within reach, the child has to correctly make one of them perform an action on the other. In the picture selection task, the child also hears a sentence and has to assign meaning to it. The decision involves checking both animations against the interpretation of the sentence and then physically pointing to the picture that matched that interpretation. Thus, whereas the picture selection task provides the correct interpretation for the child (which must be located), the act-out task allows the child to give the sentence any interpretation they like using the two toys.

[10] Forward movement was indicated by a looping background. Specifically, a few sections of the grass were slightly bulkier than the norm, and these moved across the screen both in land animations as well as in water animations, which had a grassy shore. This was done in order to indicate movement as subtly as possible, as opposed to, say, by including a tree, bush, or cloud. Unlike the grass bulks, these would have certainly interfered with gaze patterns.

The differences between the act-out and picture selection tasks have implications for children's expected performance and task comparability. While both tasks are susceptible to side biases of children, unscorable responses are expected to occur more frequently during the act-out task than the picture selection task since possible answers in the former are less controlled. For this reason a correct action demonstration during an act-out task may be seen as stronger proof of comprehension than a correct point during picture selection.

The third task that tests comprehension requires no conscious decision. The child hears a sentence and does not necessarily assign meaning to it. Longer looking times to the target animation than the distractor animation is interpreted as an indication that the child has assigned correct meaning to the sentence; longer looking times to the distractor is interpreted as an indication that the child has assigned incorrect meaning to the sentence; looking times that are not significantly different between target and distractor are interpreted as an indication that the child did not assigned meaning to the sentence or did not have a preference for either meaning. This task is very comparable to the picture selection task for two reasons. First, the same children were tested in both tasks. Second, the exact same pictures and trials were used in both tasks, but during a session on a different day. However, differences between the two tasks may be expected since previous studies have found better comprehension on a preferential looking task than on tasks that involve a conscious decision (Brandt-Kobele & Höhle, 2010; Chan et al., 2010)

Overall, what makes the three different comprehension tasks comparable is the use of the same test sentences across all three tasks. In addition, materials in each task were controlled for size saliency, agent/target side, and direction of action in a similar fashion. Since each task is designed for children between the ages of two-an-a-half and four, no significant performance differences were expected between the tasks, but rather within the tasks due to the animacy manipulation.

*Across comprehension and production*

Although the elicited production task is different in nature from the comprehension tasks, measures are taken to make the two types of tasks comparable. In the elicited production task, a child describes an action they see either performed with two toys by an experimenter or depicted in an animation on a computer screen. The child sees a single action, has to determine who was agent and who was patient, and has to describe the action. This involves finding names for two entities, conjugating a verb, and articulating a sentence. In order to make comprehension and production as comparable as possible, the same group of children are tested on the same test sentences in both tasks. That is, the children who participate in the act-out comprehension task with toys also participate in the elicitation task with toys; and the children who participate in the picture selection and preferential looking tasks with animations also participate in the elicitation task with animations. The comprehension and production tasks are further comparable since the exact same toys/target animations were used in each.

Both C&M and McClellan, Yewchuk, and Holdgrafer (1986) found that preschoolers performed better on the production of S-O word order than on comprehension. Extra-grammatical explanations for such asymmetries discussed in Section 2.2.1.1 make reference to experimental limitations (e.g. perhaps the production task is easier than the comprehension task) and cognitive limitations (e.g. perhaps the comprehension of word order requires more processing resources than production). Indeed, in the production task the child is always presented with a single meaning that must be linguistically encoded. This is in contrast to the comprehension task where the child may consider more than one possible meaning before arriving at an answer. However, in order to answer in a production task, the child must constrain him or herself to a single sentence formula, find names for the entities, and correctly order them in the articulated sentences (Levelt et al., 1999). It is for these reasons that comprehension is traditionally seen as easier than production (Bates, Dale, & Thal, 1995; Clark & Hecht, 1983b) and why an asymmetry in the opposite direction is so striking.

The term "better production" should, however, be clarified for a full picture of C&M's findings. It is not the case that the children in the C&M study performed flawlessly on their produced sentences. Rather, based on scorable responses alone it was concluded they reversed their word order in production less often than in comprehension. Figure 3.1 shows how many of the 360 total responses were correct, incorrect, or unscorable. While about 25% of the comprehension responses were unscorable, over 50% of the production responses were unscorable. This was likely the case for two reasons. First, despite efforts to prime scorable sentences, children may nonetheless produce other types of utterances (e.g. *The boat and the airplane are taking a ride*). Second, very young children may not produce full sentences at all (e.g. *boat*). The children in C&M were between the ages of 1;8 and 2;8 and most children were re-administered production items in a second or third session.

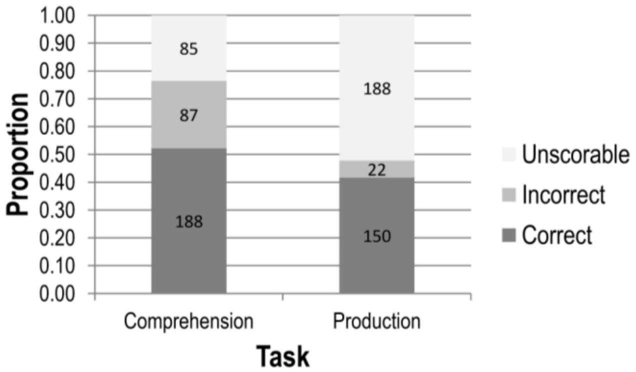


**Figure 3.1 Scorability of comprehension and production responses in Chapman & Miller (1975)**

Concluding that production of S-O word order precedes the comprehension of S-O word order based on scorable answers only is fair. It may seem misleading to ignore the fact that upwards of 50% of productions are thrown out in a study with preschoolers that claims that there is an asymmetry in favor of production. However, what is of interest is the number of OS interpretations and OS productions. If children allow more OS responses in comprehension than in production—especially in sentences that violate an animacy constraint—then this supports the claim that there is an asymmetry.

One final note on the comparison of production and comprehension addresses the fact that children in the current experiments always received comprehension blocks before production blocks. If children produce word order better than they comprehend it, could it be because of this fixed ordering? Probably not. By receiving the comprehension task first, in which they hear a sentence frame repeated, the children are given the chance to understand the types of sentences sought by the experimenter in the production task. This helps keep testing sessions short and efficient, since no follow-up sessions were included in the design. Crucially, the child receives no instructive feedback if they act out a sentence or point to an animation representing an OS interpretation. So the child is not taught anything about word order during the comprehension block. This is simply the most efficient task ordering, assuming most children are sensitive to the priming: It is likely that a child would perform the same in the comprehension task, regardless of in which block they receive it, whereas the same child would produce many more unscorable responses in the production task if they received it first rather than second.

*Children vs. adults*

The experimental design and procedure was not the same for children and adults. Adults were tested on the same test sentences in both comprehension and production, but only with the picture selection and elicitation tasks with animations. Since these tasks were administered on the eye tracker, which allowed online measures of RT and gaze data to be collected, they were more appropriate as a control task than the act-out task, which only provided offline responses. The design differed in that the order in which adults receive the comprehension and production tasks is counterbalanced, the *push* and *pull* are tested in the same list, 16 fillers are added to mask the question being investigated, and answers in comprehension are given via button press rather than finger point. These differences are necessary to tailor the design to an older group of participants and are not expected to affect the results.

By testing adult controls with the same materials as those used with children, it can be determined whether there is indeed no asymmetry between comprehension and production for adults. If adults were to exhibit a proportion of SO order in production that is greater than the proportion of SO interpretations in comprehension, then the asymmetry preschoolers exhibit would not be a developmental phenomenon. The testing of adults also serves as a control of the animations used, since adults tested with the production task first had never seen the animations before having to describe them. Had the adults not been able to reliably describe an animation, it would have been removed from the analysis for children. Furthermore, the measures of RT and eye movements help determine if there is an online effect of the animacy manipulation. Such effects are argued to be evidence for a low-ranked animacy constraint (Lamers & de Hoop, 2005).

To sum up this section about experimental controls: the sentences used in this investigation isolate the effects of word order and animacy by controlling for case, agreement, and verb selectional restrictions. There are also no further confounds of event probability, animacy categories, or context. The materials used, in the form of toys or animations, prevent visual confounds related to size and saliency. All of these controls allow for an optimal comparison across comprehension tasks, across comprehension and production, and across children and adults.

## 3.4 Summary and overview of predictions

This chapter has given an overview of methods for testing comprehension and production in preschoolers, and it has described aspects of the current experimental design that controlled for confounding factors. Here I provide a brief summary of each section, together with general expectations that integrate the predictions of the theoretical models discussed in Chapter 2 with the methodology discussed in the present chapter.

The first section of this chapter addressed methods of testing production in preschoolers. Sentence elicitation is a task in which an experimental participant witnesses an event and describes it linguistically. This type of data collection is more appropriate for the present investigation than the collection of spontaneous speech since we have knowledge about and control over the events being described. Because participants have the freedom to describe events however they wish, responses may be elicited in which S-O word order cannot be determined. By giving the production task after the comprehension task to children, the likelihood of eliciting sentences with S-O word order (whether it be SO or OS) is increased.

Based on Hendriks, de Hoop, and Lamers' (2005) model of word order development, both adult and preschool-aged speakers of English and Dutch will use SO word order in production. The elicitation task in this study measures the proportion of SO and OS word order used by participants when describing an event that is acted out before them or depicted in a cartoon animation. It is expected that there will be a high proportion of SO word order used by children and adults, across all four sentence types. Gaze measured during sentence elicitation by children and adults is expected to reflect a search for agent and then patient, as was found by Griffin and Bock (2000) for adult speakers of English. Although word order is not expected to be affected by animacy, the agent-then-patient gaze pattern and speed of response measurement (VOL) may be facilitated by animacy, to the extent that animate entities are inherently more accessible than inanimate entities (e.g. van Bergen, 2011: 20-25).

The second section of this chapter addressed methods of testing comprehension in preschoolers, focusing on three relevant tasks. In an act-out task, a participant hears a sentence and acts it out with toys. In a picture selection task, a participant hears a sentence and selects a matching picture from a set of choices. In a preferential looking task, a participant is presented with a set of choices and simply looks; his or her gaze towards each choice is measured to create a time course of eye movements. All three tasks are appropriate for preschool-aged children, but they differ in a few ways. The act-out task is more prone to unscorable responses than the picture selection task. The preferential looking task is often believed to be the easiest task that reflects the greatest amount of linguistic competence.

Based on Hendriks et al.'s model of word order development, preschool-aged speakers of English and Dutch will have SO interpretations least often for S<O sentences compared to the other three sentence types. I further predict that children will have SO interpretations most often for S>O sentences, with S=O sentences settling somewhere between—due to either gradual constraint re-ranking, an incremental application of constraints, or both. The proportion of SO and OS interpretations children have is measured in several ways in the present study, reflected by their performed actions, their animation selections, or their looking behavior. If children exhibit the predicted pattern on multiple tasks, this would be seen as strong evidence supporting the presence of a too highly ranked animacy constraint in their grammar.

Because the word order constraint is ranked above the animacy constraint in the grammars of adult speakers of English and Dutch, they are expected to have SO interpretations for all four sentence types. Based on my model implementing de Hoop and Lamers' (2006) incremental constraint optimization, however, some processing effects of the animacy constraint are expected. Namely, sentences with inanimate subjects, especially S<O sentences, are predicted to be at a disadvantage during processing (these predictions are given in Tableau 2.8). Gaze patterns reflect speed of processing during online sentence comprehension; reaction time gives an indication of the total amount of processing time necessary to come to an interpretation. These measurements collected for adults are expected to reflect a facilitation of [+an -an] sentences (shorter RTs and greater degree of looking to target), as well as a difficulty with [-an -an] and [-an +an] sentences (longer RTs and slower looks to target).

The third section of the chapter addressed the importance of experimental controls. It was pointed out that in the everyday input children receive, animacy often works together with other sources of information such as case, agreement, and verb selectional properties. In order to test the hypothesis that animacy will interfere with word order information in comprehension, all of these other sources of information must necessarily be stripped away. Reasons were given for why event probability, broad animacy categories, context, and visual saliency are potentially confounding in a study of S-O word order. Because much care has been taken in developing the experimental design in the present study, results are able to be reliably compared across comprehension tasks, across adults and children, and crucially, across comprehension and production. Thus, it is presumed that the experimental design elaborated on in this chapter has been sufficiently controlled to allow conclusions to safely be drawn concerning the two main questions at hand: is SO order used by preschoolers in production more often than it is interpreted as such in comprehension, and does animacy influence the comprehension of word order?

# 4 Comprehension: Act-out with toys

Experiments 1a & 2a

This chapter presents two experiments carried out with Dutch preschoolers. In each experiment, comprehension of S-O word order is tested with an act-out task. The accompanying sentence elicitation experiments carried out with the same populations are presented in the next chapter. (An overview of all experiments is provided in Table 1.1 at the end of Chapter 1.)

## 4.1 Introduction

Several studies that tested word order comprehension of English-speaking children with an act-out task were presented in Section 2.1.1 and summarized in Table 2.1 (Akhtar & Tomasello, 1997; Bates et al., 1984; Chan et al., 2009, 2010; Chapman & Miller, 1975; Childers & Tomasello, 2001; McClellan et al., 1986; Strohner & Nelson, 1974; Thal & Flores, 2001; de Villiers & de Villiers, 1973). The general finding among the studies is that English-speaking children approximately between the ages of 2;6 and 3;0 exhibit variable comprehension of word order in simple, reversible sentences. In general, children interpret word order best when the subject is animate and the object is inanimate (S>O), worst when the subject is inanimate and the object is animate (S<O); and roughly in between when the subject and object are equal in animacy (S=O). German children in Chan et al. (2009) also exhibited this pattern—not only at age 2;6 but also at age 3;6, when English-speaking children have already come to reliably use word order to interpret sentences.

The question that Experiments 1a and 2a aim to answer is whether this pattern of variable comprehension is found in a well-controlled act-out task carried out with Dutch preschoolers. Hendriks, de Hoop, and Lamers' (2005) model of early S-O word order predicts variable comprehension in Dutch preschoolers, whose grammar initially gives too much priority to an animacy constraint on meaning. Because Dutch and German children receive similar input with regard to word order, I have predicted that variable comprehension will persist in Dutch children until age 3;6.

## 4.2 Experiment 1a

### 4.2.1 Method

*Participants*

Monolingual Dutch 2½-year-olds ($n$ = 14, 9 male, range 2;5 – 3;2, *mean* = 2;9, *sd* = 3.2 mo.s) and 3½-year-olds ($n$ = 16, 8 male, range 3;4 – 3;11, *mean* = 3;8, *sd* = 2.5 mo.s) participated in the study. The children attended day cares in the Groningen area, and signed permission was obtained from their parents on a volunteer basis. An additional 8 children were invited to participate: 7 (aged 2;9 or younger) were either unresponsive or unable to perform the task with simple practice items, and the remaining child (3;3) was able to be tested, but was discovered afterwards to be bilingual.

Information about each child's language development was collected by asking parents to fill in one of three vocabulary checklists. The Dutch ("N") version of the MacArthur-Bates Communicative Development Inventory (CDI) II and III short forms were completed by parents with children between the ages 2;0 and 2;6 (N-CDI II) and between 2;6 and 3;2 (N-CDI III) (Fenson et al., 2000; Zink & Lejaegere, 2003, 2007). The Dutch versions of the CDI have been normed and tested for validity, which allows each child's score to reliably place him or her in a percentile based on age—and in the case of the N-CDI II, also on gender. The parents of children aged 3;3 or older filled out the vocabulary section of the *Kleuter Inventarisatie Nederlandse Taalverwerving* (KINT) language inventory (Koster, Plas, & Krikhaar, 2004). The KINT is designed in line with the principles and set-up of the N-CDI, but is intended for children between the ages of three and five. While the KINT has not yet been normed, it provides information about how the older children rank among themselves with regard to vocabulary development.[11]

*Design and materials*

The 16 test sentences were discussed in detail in Section 3.3.1 and appear in Table A.2 in Appendix A. They consist of 8 *duwen* (*push*) and 8 *trekken* (*pull*) sentences. Each of the following conditions occur twice for each verb: animal subject with animal object [+an +an], animal subject with vehicle object [+an -an], vehicle subject with animal object [-an +an], and vehicle subject with vehicle object [-an -an]. Sentences in List B are the same as those in List A, but with the two NPs reversed. In addition to these test sentences, there were 6 practice items and 4 filler items (listed in Table A.4). All experimental sentences were pre-recorded and spoken with neutral prosody by a female voice.

The toys used in the experiment were 14 bath toys and 1 plastic block. Of these 15 toys, 8 were used in the test sentences (shown in Figure B.1 in Appendix B) and the remaining 7 were used in the practice and filler items (Figure B.2). All toys were approximately the same size, each measuring about 7 cm x 4 cm x 6 cm. A square piece of soft Velcro was attached to the bottom of each of the test item toys and four of the practice/filler item toys. A square piece of hard Velcro was sewn to each end of a 15 cm long piece of sisal twine, or rope-like string. Two such ropes were created to enable pulling.

Items were prepared so that *agent side* was balanced: the toy corresponding to the agent in the spoken sentence was placed to the right of the patient half of the time, and to the left of the patient half of the time, across sentence types. Furthermore, in the order of presentation, adjacent items shared no nouns nor were they of the same sentence type.

*Procedure*

The researcher and native Dutch-speaking assistant invited a child to a quiet room at the day care to sit at a table across from them. Before the act-out task was begun, two pre-tests were carried out. The child was first asked to name each of the 15 toys. If the child did not know a word or used a different word, the assistant gave feedback and made sure to present the toy again so the child was familiar with it.[12] The

[11] Age groups are defined by the cut-off age 3;2 / 3;3 (rather than 3;0 / 3;1) because the vocabulary list available for the older group (KINT) provides no percentile scores. Percentile scores are based on norms established by a larger population of Dutch children and allow children to be ranked according to their vocabulary abilities *across* different vocabulary lists. Since this was not possible, the cutoff age for the younger group was raised by two months so that *vocabulary score* could be included as a potential predictive factor of performance in the analysis. The younger children were ranked based on percentile score (from the N-CDI norms), and the older children were ranked based on their raw scores.

[12] Two toys stood out as tricky for the children: *helikopter*, and *brommer* (*moped*), the latter of which they often preferred to call a *motor* (*motorcycle*). Because sentences were pre-recorded, it was not the case that whichever word the child preferred (*brommer* vs. *motor*) could be used. Ultimately, this was not problematic since the children were made familiar with the target terminology and these toys appeared in filler and practice items only.

second pre-test involved a demonstration of the actions of pushing and pulling. The moped was linked to the helicopter with a piece of rope, and the pig was attached to the horse in the same way (see Figure B.3). The assistant first used the helicopter to push the moped and said "Look, this is 'pushing.' Now you try." The assistant indicated to the child to demonstrate *pushing* using the two animals, while very carefully avoiding any modeling of S-O word order. Next, the assistant used the pig to pull the horse and said, "Look, this is 'pulling.'" As the child successfully used the moped to pull the helicopter, the assistant repeated "Yes, that's 'pulling', very good!" If a child had difficulty demonstrating either verb, the assistant used different strategies to guide the child until the child succeeded.

The child was then tested in two blocks with a small break in between. In each block, only one of the test verbs was tested. Whether *push* or *pull* was tested in the first block was balanced over list (A or B), gender, and age group. The task, described below, was preceded by three practice items in the first block and by two practice items in the second block. During the practice items, the child was given feedback to indicate whether he or she had successfully played the game by paying attention to the sentence (rather than, for example, just playing with the toys). The action was not modeled for the child, nor was negative feedback given if the child exhibited a reverse interpretation of S-O word order. Each block also included two filler items to keep the task interesting.

The researcher placed two toys about 10 cm apart on a tray in front of the assistant. The assistant sat directly across from the child, and the toys were placed facing the child. For *pull* items, the rope connected the two toys: the rope jutted out towards the front of each toy so that it was visible to the child and so that it did not indicate which item should be the puller/pullee. The assistant then played a sound file on a laptop computer that contained the test sentence, an instruction, and a repetition of the test sentence: *De auto duwt de koe. Laat het maar zien. De auto duwt de koe.* There was a gap of 500 ms between each sentence. The first test sentence was identical to the second instance of the test sentence. The *Laat het maar zien* translates roughly as *Show it* or *Let's see it*. It was during this middle part of the sound file that the assistant placed the tray in front of the child. If the child was unresponsive, performed an unclear action, or performed an action other than pushing or pulling, the assistant repeated the sentence with neutral prosody. In the end, the child was praised for whatever action was performed in order to keep the child motivated.

During the break the child was able to leave the table for a few minutes, play with a toy, and drink some juice. At the end of the session the child was offered a small toy before being brought back to the regular play area of the day care.

*Scoring*

Responses were categorized as scorable (*SO* or *OS interpretation*) or unscorable (*Unclear* or *Other action*). If a revision was made by the child, the final action was used for scoring. Actions clearly corresponding to a SO interpretation were marked as *SO interpretation*, and actions clearly corresponding to an OS interpretation were marked as *OS interpretation.* If a child consistently used one toy to knock the second toy over during the *push* sentences, this was accepted as a *push* action. If it was unclear from a series of actions whether a response was a *SO* or an *OS interpretation*, the action was marked as *Unclear* (for example if the child pulled the two toys back and forth repeatedly). Several examples of a response marked as *Other action* are if the child: picked one or both of the toys up, waved them around, and placed them back down; placed one toy on top of the other; made one toy kiss or jump over the other; smashed one toy into the other repeatedly; switched the location of the two toys; knocked one or both of the toys over with hands; or caused the toys to growl at each other. A second categorization of responses was conducted by an independent scorer; agreement was high (Cohen's $\kappa = .81$).

It was possible that despite encouragement, the child would give no response to an item at all during one of the tasks, in which case the item was scored as *No answer*. In the case that a child wanted to stop the game, the test was ended, rendering the remaining missing items for the task in that block as *Not given*.

### 4.2.2 Results

*Criteria for inclusion*

A child was included in the analysis if he or she contributed at least two scorable responses per sentence type. Two children were excluded because they had too many unscorable or missing responses. Their removal decreased the number of 2½-year-olds by 2 (*n* = 12, *mean* = 2;9, *sd* = 3.1 mo.s). The number of 3½-year-olds did not change (*n* = 16, *mean* = 3;8, *sd* = 2.5 mo.s).

*Scorability*

The actions performed by the children were largely scorable. Of the 191 responses given by the 2½-year-olds, 177 items were scorable and 14 were unscorable (*Other action*). (One item was *Not given*). Of the 255 responses given by the 3½-year-olds, 252 were scorable and 3 were unscorable (*Other action*). (One item received *No answer*.)

*Accuracy*

Mean accuracy (*SO interpretations*) for each age group on the four sentence types based on participant means are shown in Table 4.1 and Figure 4.1. On average, the 2½-year-old's responses reflected *SO interpretations* 59% of the time (and *OS interpretations* 41%), while the 3½-year-olds's responses reflected *SO interpretations* 92% of the time (and *OS interpretations* 9%). (Individual scores can be found in Table C.1 and Table C.2. in Appendix C.)

Multilevel regression modeling is used to determine what effects the manipulated variables have on the measured outcome (here, accuracy). Multilevel modeling has advantages over more traditional analyses in that it simultaneously accounts for between and within participant behavior, and it corrects for differences in sample size between participants (e.g. Baayen, 2008: 260-266; Gelman & Hill, 2007: 246). Models can be corrected for differences between participants and between items by including these as random factors. By doing so, significant main effects or interactions can be generalized beyond the particular participants and items used in the experiment (e.g. Baayen, Davidson, & Bates, 2008). Furthermore, multilevel modeling can be used with binomial data (such as SO vs. OS interpretations measured here) without transforming the data; instead, the model is adjusted to expect that the variance increases with the mean according to the binomial distribution (Baayen, 2008: 197).

**Table 4.1 Mean proportion SO interpretations: Experiment 1a**

| Sentence type | 2½yrs | *sd* | 3½yrs | *sd* |
|---|---|---|---|---|
| + animate + animate | .59 | *.26* | .97 | *.09* |
| + animate - animate | .63 | *.35* | .93 | *.15* |
| - animate + animate | .60 | *.36* | .92 | *.13* |
| - animate - animate | .53 | *.34* | .87 | *.23* |
| Total | .59 | *.12* | .92 | *.11* |

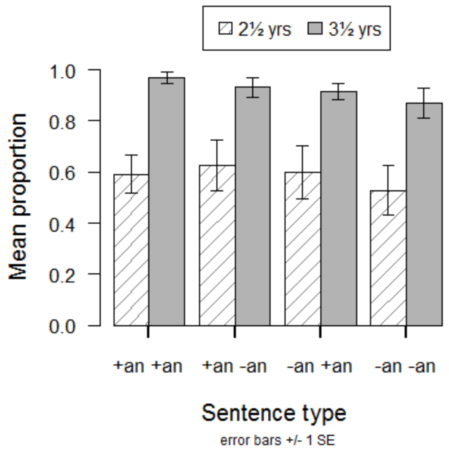


**Figure 4.1 Mean proportion SO interpretations: Experiment 1a**

By comparing different models using a simplification procedure, a model with the best fit can be determined. First, a complete interaction model is created. For example, a model is created that includes a three-way interaction of the factors subject animacy, object animacy, and age group, all possible two-way interactions between these factors, and the main effects of each of the factors. Then this complete interaction model is compared to a simpler model without the three-way interaction using a chi-square test that evaluates each model's goodness of fit given the degree of freedom (Baayen, 2008: 253-255). If a simpler model has a significantly lower goodness of fit than a more complex model, then the deletion of a factor or interaction is not justified. This model comparison process is repeated until the best model has been determined. The final model is either (i) a baseline model with no interactions of the factors (e.g. subject animacy, object animacy, and age group), each of which may or may not contribute main effects; or (ii) a model with the baseline factors as well as interactions that have been shown to significantly explain variance in the data. By reporting the beta coefficients of the final model's significant factors, the direction of each significant effect can be determined. Thus, a $\chi^2$-value is reported to show whether the inclusion of a factor is or is not necessary to explain variance in a model, and a $\beta$-value of a significant factor is reported to show the direction of that factor's effect in the final model. The $p$-values of the two statistics for a factor are usually the same or very close.[13]

To determine whether S-O animacy affected accuracy in comprehension in either age group, the binomial data were fit to a mixed effects model with subject animacy, object animacy, and age group as fixed factors, and participants and items as random factors. There was no significant three-way ($\chi^2(1)$ = .36, $p > .1$) or two-way ($\chi^2(3) = 3.21$, $p > .1$) interaction between the fixed predictors. Since including interactions was not justified, the baseline model was checked for main effects. There was no main effect of subject animacy ($\chi^2(1) = 1.07$, $p > .1$) or object animacy ($\chi^2(1) = 1.07$, $p > .1$). Age group, however, was a significant predictor of accuracy ($\chi^2(1) = 27.10$, $p < .001$), with 3½-year-olds performing better than 2½-year-olds ($\beta = 1.15$; $z = 6.16$; $p < .001$). The inclusion of control factors such as gender, test verb, agent side, list, and vocabulary score did not significantly explain more variance in the data. Thus, the 3½-year-

[13] The multilevel regression was carried out using the free software package R (R Development Core Team, 2011). The *lmer* function in package lme4 was used to obtain coefficient estimates for all data and additionally the $p$-values for binary data ($z$-statistic is reported) (Bates, 2007). The *pvals.fnc* function in package languageR was necessary to obtain $p$-values for continuous data using a Markov Chain Monte Carlo method ($t$-statistic is reported) (Baayen, 2008; Baayen et al., 2008).

olds responded with more actions reflecting SO interpretations than the 2½-year-olds, and the animacy manipulation had no significant effect on either group.

### 4.2.3 Discussion

The results of Experiment 1a show that children's responses in the act-out task did not vary as a function of the manipulation of subject and object animacy, neither at age 2;6 nor 3;6. The only reliable predictor was age, with the older preschoolers performing better than the younger. The older children's responses reflected a SO interpretation of word order 92% of the time, whereas the younger children's responses reflected a SO interpretation 59% of the time.

It is not clear from these results what factors, if not subject or object animacy, are preventing the younger children from using word order to interpret the sentences. It is possible that the distance in animacy between animals and vehicles was too small to affect the interpretations of children in either age group. In order to see whether greater difference in animacy between the animate and inanimate nouns has an effect on sentence comprehension a follow-up act-out experiment was carried out (Experiment 2a) in which the animals were replaced by humans.

## 4.3 Experiment 2a

### 4.3.1 Method

*Participants*

Monolingual Dutch 2½-year-olds (*n* = 32, 17 male, range 2;6 – 3;2, *mean* = 2;10, *sd* = 2.4 mo.s) and 3½-year-olds (*n* = 33, 18 male, range 3;3 – 3;10, *mean* = 3;6, *sd* = 2.2 mo.s) participated in the study. The day cares granted permission to test the children, and parents could opt not to have their child participate. An additional 5 children were invited to participate (ranging from 2:11 to 3;10) but they were either unresponsive, unable to perform the task with simple practice items, or unwilling to continue the game after a few test items. No vocabulary information was collected.

*Design and materials*

The same experimental design was used as in Experiment 1a, but the following changes were made to the materials. Each test sentence animal was replaced by a human: *koe*, *hond*, *eend*, and *vis* (*cow*, *dog*, *duck*, and *fish*) were replaced by *mevrouw*, *meisje*, *meneer*, and *jongen* (*lady*, *girl*, *gentleman*, *boy*), respectively. All animal bath toys used in Experiment 1a—also those used in the practice and filler items—were replaced by LEGO DUPLO® toys. Pictures of the toys are found in Appendix B, Figures B.4 – B.6. The gentleman and lady have grey hair, and the boy and girl have dark hair.[14] The sisal twine was replaced by soft yellow string to prevent the new, lighter figures from falling over. New sentences were recorded (only for the adjusted sentences) by the same female speaker recorded in Experiment 1a.

---

[14] Regarding the gray-haired figures: pilot testing showed that *grandpa* and *grandma* were not easier labels than *gentleman* and *lady*. In addition, the former in Dutch, as in English, are usually used as proper rather than common nouns. Regarding the dark-haired figures: these figures representing a boy and a girl were more youthful than those representing the man and the lady. However, LEGO DUPLO figures representing an actual boy and girl were not used because they are smaller in size than the man and lady. Occupational labels (e.g. *fireman*, *doctor*) were not used since they tend to be described with longer words, usually include salient accessories like hats, and are less familiar to children than the basic words used here.

*Procedure*

The same testing procedure was used as in Experiment 1a except for an adjustment of the pre-test for the naming of human figures. First, the humans were presented together in pairs based on age/generation and the assistant said, for example, “This is the gentleman and lady.” The child was then asked to identify which was which, effectively identifying differences in gender based on the labels given. Second, the humans were presented together in pairs based on gender and were named by the assistant, for example "This is the girl and the lady". The child was then asked to point to which was which, thereby identifying differences in age based on the labels given. The pre-test was carried out in this way to prevent the child from getting confused by their own ideas about possible names for the figures. The crucial distinction the children had to make was between humans of the same gender (e.g. *girl* vs. *lady*) since these appeared together in test items.

*Scoring*

The same scoring procedure was used as in Experiment 1a. Inter-scorer agreement was high (Cohen’s κ = .88).

### 4.3.2 Results

*Criteria for inclusion*

There were 2 children excluded from the analysis because they did not contribute at least two scorable responses per sentence type. The removal of these 2 participants decreased the number of 2½-year-olds (*n* = 30, *mean* = 2;10, *sd* = 2.4 mo.s). The number of 3½-year-olds did not change (*n* = 33, *mean* = 3;8, *sd* = 2.2 mo.s).

*Scorability*

The actions performed by the children were largely scorable. Of the 480 responses given by 2½-year-olds, 444 were scorable and 36 were unscorable (*Other action)*. Of the 525 responses given by 3½-year-olds, 511 were scorable, 14 were unscorable (*Other action)*. (2 received *No answer* and 2 were *Not given.*)

*Accuracy*

Mean accuracy for each group on the four sentence types based on participant means are shown in Table 4.2 and Figure 4.2. On average, the 2½-year-olds’s responses reflected *SO interpretations* 62% of the time (and *OS interpretations* 38%), while the 3½-year-olds’s responses reflected *SO interpretations* 83% of the time (and *OS interpretations* 17%). (Individual scores can be found in Table C.3 and Table C.4. in Appendix C.)

**Table 4.2 Mean proportion SO interpretation: Experiment 2a**

| Sentence type | 2½yrs | *sd* | 3½yrs | *sd* |
|---|---|---|---|---|
| + animate + animate | .62 | *.27* | .81 | *22* |
| + animate - animate | .62 | *.30* | .88 | *20* |
| - animate + animate | .64 | *.31* | .76 | *33* |
| - animate - animate | .62 | *.26* | .85 | *21* |
| Total | .62 | *.14* | .83 | *17* |

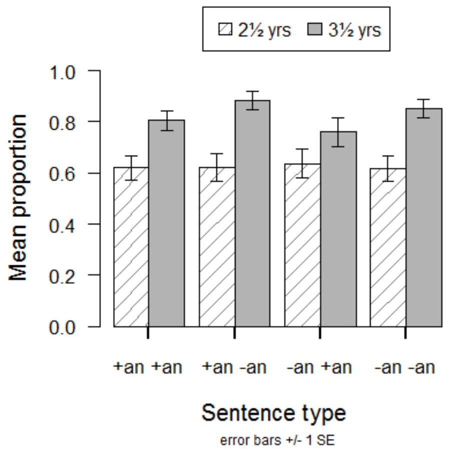


**Figure 4.2 Mean proportion SO interpretation: Experiment 2a**

To determine whether S-O animacy affected accuracy in comprehension in either age group, the binomial data were fit to a model with subject animacy, object animacy, and age group as fixed factors, and participants and items as random factors. There was no significant three-way interaction between the fixed predictors ($\chi^2(1) = .21$, $p > .1$). There was also no significant two-way interaction between subject and object animacy $\chi^2(1) = .05$, $p > .1$) or age and subject animacy ($\chi^2(1) = 1.60$, $p > .1$). There was a significant interaction of age group and object animacy ($\chi^2(1) = 5.79$, $p = .02$) as well as a significant effect of age group ($\chi^2(1) = 23.50$, $p < .001$). This means that the 3½-year-olds responded with more actions reflecting SO interpretations than the 2½-year-olds in general ($\beta = .64$; $z = 5.17$; $p < .001$), but they did so to a significantly lesser degree when the object was animate ($\beta = -.20$; $z = -2.41$; $p = .02$).

The inclusion of control factors such as test verb, agent side, and list did not significantly explain more variance in the data. However, the inclusion of the interaction of gender and subject animacy in the model above made a contribution that approached significance ($\chi^2(1) = 2.50$, $p = .11$), indicating that males tended to perform better than the females when the subject was inanimate ($\beta = .13$; $z = 1.60$; $p = .11$). Children's mean performance on each sentence type based on gender and age group are shown in Table 4.3. Thus, while there was no significant interaction between gender and subject animacy, there was a non-significant trend showing that male participants more readily accepted vehicles as subjects than females did.

**Table 4.3 Mean proportion SO interpretation by gender: Experiment 2a**

| Sentence type | 2½yrs Male (*n* = 16) | *sd* | 2½yrs Female (*n* = 14) | *sd* | 3½yrs Male (*n* = 18) | *sd* | 3½yrs Female (*n* = 15) | *sd* |
|---|---|---|---|---|---|---|---|---|
| +an +an | .61 | *.28* | .63 | *.25* | .79 | *.25* | .82 | *.19* |
| +an -an | .57 | *.29* | .67 | *.30* | .89 | *.20* | .88 | *.21* |
| -an +an | .68 | *.33* | .58 | *.30* | .80 | *.28* | .71 | *.39* |
| -an -an | .67 | *.25* | .56 | *.27* | .82 | *.22* | .89 | *.19* |

### 4.3.3 Discussion

The results of Experiment 2a show that children's responses in the act-out task did not vary as a function of the manipulation of subject and object animacy at age 2;6. However, it did affect the children's performance at age 3;6. The younger children's responses in Experiment 2a reflected a SO interpretation of word order 62% of the time. The older children's responses reflected SO interpretations more often than the younger children (83% of the time), but this effect of age was not as strong for the two sentence types in which the object was animate.

The greater distance in animacy between the humans and vehicles in this experiment compared to the distance between the animals and vehicles in Experiment 1a may explain why the 3½-year-olds disprefered the animate entities as objects. In other words, the greater animacy of humans over animals may explain why the group of children in Experiment 2a tested with sentences like *The car is pushing the lady* were more likely to interpret them as OS than the group of children tested in 1a with sentences like *The car is pushing the cow*. When considering [-an +an] sentences, the effect of object animacy in the second experiment is in line with the account of Hendriks et al. (2005), whose OT model of S-O word order acquisition predicts that a SO interpretation is optimal for S>O and S=O sentences, but not S<O sentences. When considering the [+an +an] sentences (S=O), the effect of object animacy may be explained by the fact that, under certain formulations of the animacy constraint (outlined in Section 2.3.1.2), it is not necessarily the relative animacy of subject and object that guides interpretation, but the individual preferences for animate subjects and inanimate objects. An alternative explanation for the performance on the [+an +an] sentences is that children occasionally confused the names for the human toys used in the experiment, and the [+an +an] condition was the only in which the names of the human toys necessarily had to be distinguished.

The prediction of the OT model was not met by the performance of the 2½-year-olds, even with the greater distance in animacy between subject and object in Experiment 2a. Possibly related to this finding is the observation that the male children tended to accept vehicles as subjects more readily than their female counterparts did. While it could be the case that the 2½-year-olds were simply not influenced by an animacy constraint, a second possibility is that effects of the animacy constraint were obscured by a confound related to the use of toy vehicles during testing with young boys. While gender was not expected to play a role in the experiments, the trend that was found is not surprising: a general observation during testing was that boys occasionally exhibited an affinity for the vehicle toys that the girls did not. The role that gender may play in the results is discussed further in the general discussion that follows.

## 4.4 General Discussion

The question that Experiments 1a and 2a aim to answer is whether Dutch preschoolers exhibit variable comprehension of word order due to S-O animacy. The results of Experiments 1a and 2a are summed up as follows:

- Experiment 1a tested the comprehension of 30 preschoolers using an act-out task with sentences in which toy animals and vehicles interacted.
  - No variation due to S-O animacy was found in the word order comprehension of the 2½- or 3½-year-olds tested.
  - Age was a significant predictor, with 3½-year-olds interpreting sentences as SO more often than younger children (92% vs. 59%).

- Experiment 2a tested the comprehension of 65 preschoolers using an act-out task with sentences in which toy humans and vehicles interacted.
    - There was no variation due to S-O animacy found in the word order comprehension of the 2½-year-olds.
    - The 3½-year-olds exhibited variable comprehension: they performed better than the 2½-year-olds in general, but to a lesser extent when the sentence objects were animate.

Thus, variable comprehension was found, but only in the older children in the second act-out task, which had a larger distance of animacy between subject and object than the first act-out task.

The variable comprehension that was found in Experiment 2a is partly in line with the predictions made. First, the OT model of Hendriks et al. (2005) which ranks an animacy constraint too highly, predicts that a SO interpretation is not optimal for S<O sentences. These [-an +an] sentence types did suffer in the performance of the older children in the second act-out task, but so did the [+an +an] sentences. It was suggested that this general effect of object animacy reflects the older children's dispreference for animate objects, which can be viewed as "half" of the predictions made by de Hoop and Lamers' animacy constraint, which stresses the importance of the relative animacy of subject and object. The effect of object animacy is certainly in line with other formulations of an animacy constraint defined as individual preferences for an animate subject and an inanimate object. Second, it was predicted that variable comprehension could be found in Dutch preschoolers of age 3;6, even though English-speaking children of the same age have come to rely on word order. This prediction was met in the second act-out task.

It is not clear why the 2½-year-olds did not exhibit the same variation in comprehension as the 3½-year-olds. Preschoolers acquiring English and German had, after all, been shown to exhibit variable comprehension in the face of animacy conflicts at the age of 2;6 (e.g. Chan et al., 2009). The only sign of variation was found in Experiment 2a upon inspection of the behavior of the male versus the female children: boys tended to have less difficulty than expected with sentences with inanimate subjects. In McClellan et al.'s (1986) study that used humans, animals, and vehicles, all children actually performed best on the [-an +an] sentences. This raises the issue of how individual animacy hierarchies might affect the application of an animacy constraint. It could be the case that some children view vehicles as high on the animacy hierarchy as animals or people. Bowerman (1974: 56, 84), for example, found that words for vehicles functioned like animate nouns in the speech of one young Finnish boy. It was indeed evident that some boys in both Experiments 1a and 2a were inclined to find vehicles interesting. For example, some wore shirts or hats featuring anthropomorphized car characters from popular animated films. Occasionally younger boys played with the vehicles presented in practice items before understanding the task. However, it could also be that toy vehicles are simply very salient to these boys, rather than that vehicles are too highly ranked in their animacy hierarchies. The former is likely to be the case, as no trace of gender effects are found in the experiments with preschoolers presented in Chapter 6, which use a paradigm without toys.

The fact that variable performance was found with sentences with humans and vehicles and not with sentences with animals and vehicles suggests that the animacy contrast in the latter is too weak to result in significant animacy effects. However, before concluding that an animal-vehicle contrast does not result in variable comprehension, we should consult the results of the picture selection and preferential looking experiments testing these sentences in Chapter 6. First, we turn to how the same two groups of preschoolers in Experiments 1a and 2a fared in an accompanying production task with the same materials.

# 5 Production: Elicitation with toys

Experiments 1b & 2b

This chapter presents two experiments carried out with Dutch preschoolers. In each experiment, production of S-O word order is tested with a sentence elicitation task. The accompanying act-out experiments carried out with the same populations were presented in the previous chapter. (An overview of all experiments is provided in Table 1.1 at the end of Chapter 1.)

## 5.1 Introduction

A few studies have tested word order production of English-speaking children with the same types of sentences used in Chapter 4 to test comprehension. These studies, which used a sentence elicitation task, were presented in Section 2.1.2 (Angiolillo & Goldin-Meadow, 1982; Chapman & Miller, 1975; McClellan et al., 1986). In general, English-speaking preschoolers are able to produce word order in simple sentences to describe actions that are reversible, even if the two participants in the actions are not equal in animacy.

The first question that Experiments 1b and 2b aim to answer is whether the same adult-like production of S-O word order is found in a well-controlled sentence elicitation task with toys carried out with Dutch preschoolers. Hendriks, de Hoop, and Lamers' (2005) model of S-O word order in preschoolers predicts adult-like production of word order in Dutch preschoolers. If children's production of word order is better than their comprehension of word order, this would suggest that there is a developmental asymmetry.

The second question these experiments aim to answer is whether animacy affects production. Although Hendriks at al.'s account proposes that the early grammar initially gives too much priority to an animacy constraint, the animacy constraint is one that affects meaning only. Therefore, the mis-ranking does not affect production of word order, and no variable production due to animacy is predicted.

## 5.2 Experiment 1b

### 5.2.1 Method

*Participants*

The same Dutch 2½-year-olds ($n$ = 14) and 3½-year-olds ($n$ = 16) who participated in the act-out task in Experiment 1a also participated in an elicitation task.

*Design and materials*

The same sentences and toys were used as in Experiment 1a. The direction of action was controlled for across the four sentence types. In the order in which the actions were demonstrated, no toy appeared in the preceding or following item, nor were adjacent items of the same sentence type. No filler items were included.

*Procedure*

The production procedure immediately followed the comprehension procedure described in Experiment 1a (which had been in turn preceded by a toy naming and verb demonstration pre-test). In each block, only one of the test verbs was tested. Thus, the complete session included a comprehension and production task with one verb, a small break, and a comprehension and production task with the other verb. The production task, described below, was preceded by three practice items in the first block and two practice items in the second block.

To begin, a hand puppet was introduced: Joris the Sheep. The assistant explained to the child that Joris would close his eyes so that he could not see anything. It was the child's task to tell Joris what was happening on the tray. The researcher placed two toys on the tray in front of the assistant so that they were facing the impending direction of action. In the case of *pull* sentences, the toys were connected with rope. The assistant then performed an action and encouraged the child to describe it.[15] The phrase, "Wat gebeurt er?" ("What's happening?") was used instead of "Wat zie je?" ("What do you see?") in order to encourage an answer that described the entire action. The assistant continued to perform the action for the child until the child uttered a sentence or until it was clear that the child would not answer or would only name the toys or action. Joris then opened his eyes and check to see if the child was right. Joris always gave positive praise except in the case of no answer, in which case he would encourage the child to try again on the next item. The same small break and toy reward described in Experiment 1a was shared by the elicitation experiment.

*Scoring*

The elicited utterances were transcribed from the video recordings of the sessions. A second independent transcription of 10% of the participants ($n = 3$) showed 90% agreement.

Utterances were categorized as scorable (*SO order* or *OS order)*, or unscorable (*Nontarget*, *Insufficient*, or *Unclear*). If a revision was made by the child, the final utterance was used for scoring. *SO order* sentences contained SO word order and were either identical to the target sentence or differed in some way that did not affect the SO word order. Examples of the latter case were: a preposed subject was used (*The car, it pushes the cow*); a prepositional phrase or adverb was included (*The car pushes on the cow*); a synonym for the noun or verb was used; *gaan* (*to go*) was used as an auxiliary; or indefinite articles were used or articles were omitted. *OS order* sentences contained OS word order and were either identical to the reverse of the target sentence or differed in some way that did not affect the OS word order. Unscorable responses included three different types of utterances: (1) *Nontarget* responses contained a non-synonym verb, as well as responses in which the wrong target verb was used; (2) *Insufficient* responses were those in which only the nouns were named or the action was named; and (3) *Unclear* responses referred to the toys and the correct target verb, but the word order could not be scored, for example, because of the use of a preposition like *with* (*The car pushes with the cow*), unrecoverable

[15] During the production task, it was important to keep the toys out of reach of the children, some of whom reached tirelessly across the table. Since each child was allowed to touch the toys during the comprehension task, the production task understandably brought on frustration in some. In desperate cases, the assistant allowed the child to perform the action that the assistant had been performing, if this encouraged the child to describe it.

pronouns (*He's pushing him*), or a nonfinite verb was used with one noun.[16] A second categorization of the utterances was conducted by an independent scorer; agreement was high (Cohen's κ = .91)

If the child gave no response, the item was scored as *No answer*. If the child wanted to stop the game or simply did not produce any description after several chances, the test was ended, rendering the remaining items for the task in that block as *Not given*.

### 5.2.2 Results

*Criteria for inclusion*

There were 12 children excluded from the analysis because they did not contribute at least two scorable responses per sentence type. (Two of these children had been excluded from the analysis in 1a for the same reason). The removal of these participants decreased the number of 2½-year-olds by 9 (*n* = 5, *mean* = 2;9, *sd* = 3.1 mo.s). The number of 3½-year-olds decreased by 3 (*n* = 13, *mean* = 3;8, *sd* = 2.3 mo.s).

*Scorability*

The utterances produced by the remaining children were largely scorable. Of the 75 responses given by 2½-year-olds, 63 were scorable and 12 were unscorable*:* 10 *Nontarget* and 2 *Unclear*. (1 received *No answer* and 5 were *Not given*.) Of the 208 responses given by 3½-year-olds, 207 were scorable and 1 was unscorable (*Insufficient*). Examples of unscorable utterances falling into each of the three categories are found in Table 5.1.

**Table 5.1 Examples of unscorable utterances in Experiment 1b**

| | Utterance[1] | Target sentence |
|---|---|---|
| Nontarget | *hij gaat rijden, moet op het spoor, auto mag niet*<br>it's driving, must be on the tracks, the car isn't allowed | *de trein trekt de auto*<br>the train is pulling the car |
| | *auto rijden, duwen met de koetje*<br>car driving, pushing with the cow | *de koe trekt de auto*<br>the cow is pulling the car |
| | *zwemmen in het water*<br>swimming in the water | *de eend trekt de vis*<br>the duck is pulling the fish |
| | *deze gaat poepen, wieletjes*<br>this one will poop, little wheels | *de trein duwt de hond*<br>the train is pushing the dog |
| Insufficient | *visje, een eendje*<br>fishy, a ducky | *de eend trekt de vis*<br>the duck is pulling the fish |
| | *duwen, duwen*<br>pushing, pushing | *the auto duwt de koe*<br>the car is pushing the cow |
| | *trein, en een hond, trein, hond, lopen*<br>train, and a dog, train, dog, walking | *de hond trekt de trein*<br>the dog is pulling the train |
| | *de vliegtuig met de eend*<br>the airplane with the duck | *het vliegtuig trekt de eend*<br>the airplane is pulling the fish |

[16] Constructions with nonfinite verbs are not ambiguous in English, but they are in the V2 language of Dutch. Although a noun + nonfinite verb construction could feasibly be interpreted as $(SV_{FIN})OV_{INF}$ in Dutch (e.g. "paardje rijden" (horsie to-ride) or "boekje lezen" (little-book to-read) (Neeleman & Weerman, 1997: 144)), these constructions were conservatively interpreted as *Unclear*.

| | | |
|---|---|---|
| Unscorable | *duwen met de eendje de vis*<br>pushing with the duck the fish | *de eend duwt de vis*<br>the duck is pushing the fish |
| | *de hond duwen*<br>the dog push | *de trein duwt de hond*<br>the train is pushing the dog |
| | *die duwen, die duwen*<br>that push, that push | *de koe duwt de auto*<br>the cow is pushing the car |
| | *hij gaat duwen naar deze kant*<br>he is pushing him to this side | *de koe duwt de auto*<br>the cow is pushing the car |

1.These examples come from all children in Experiment 1b, regardless of whether they were included in the production analysis

*Accuracy*

Mean accuracy (*SO order*) for each group on the four sentence types based on participant means are shown in Table 5.2 and Figure 5.1.On average, the 2½-year-olds used *SO order* 85% of the time (and *OS order* 15%), while the 3½-year-olds used *SO order* 99.5% of the time (and *OS order* 0.5%). (Individual scores can be found in Table C.5 and Table C.6. in Appendix C.)

**Table 5.2 Mean proportion SO order produced: Experiment 1b**

| Sentence type | 2½yrs | *sd* | 3½yrs | *sd* |
|---|---|---|---|---|
| + animate + animate | .85 | *.22* | 1.00 | - |
| + animate - animate | .90 | *.14* | 1.00 | - |
| - animate + animate | .95 | *.11* | 1.00 | - |
| - animate - animate | .70 | *.27* | .98 | *.07* |
| Total | .85 | *.11* | .995 | *.02* |

.

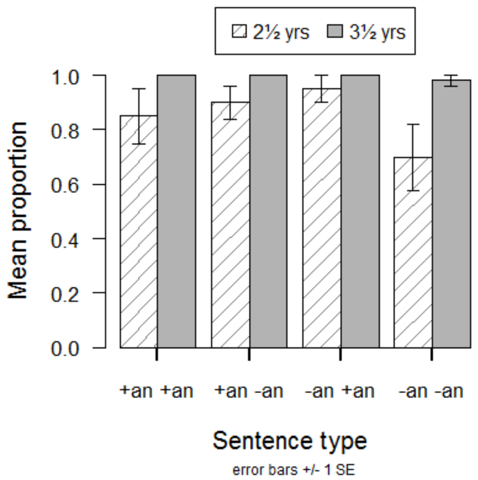


**Figure 5.1 Mean proportion SO order produced: Experiment 1b**

To determine whether S-O animacy affected accuracy in production in either age group, the binomial data were fit to a model with subject animacy, object animacy, and age group as fixed factors and participants and items as random factors. Age group x subject animacy as well as age group x object animacy proved to be collinear predictors, meaning that each of these interactions was correlated with one or both of the individual terms in the interaction. As a simple strategy to reduce collinearity, these two interactions were excluded from the models (Baayen, 2008: 183). There was no significant two-way interaction between subject and object animacy ($\chi^2(1) = 2.57$, $p > .1$). In the baseline model, there was no main effect of subject animacy ($\chi^2(1) = .22$, $p > .1$) or object animacy ($\chi^2(1) = 1.06$, $p > .1$). Age group, however, was a significant predictor of accuracy ($\chi^2(1) = 13.87$, $p < .001$), with the older children producing SO word order more often than the younger children ($\beta = 1.85$; $z = 3.47$; $p < .001$). The inclusion of control factors such as gender, test verb, direction of action, list, and vocabulary score did not significantly explain more variance in the data. Thus, the 3½-year-olds used SO order in their utterances more often than the 2½-year-olds, and the animacy manipulation had no significant effect on either group.

### 5.2.3 Discussion

The results of Experiment 1b show that children's responses in the elicitation task did not vary as a function of the manipulation of subject and object animacy, neither at age 2;6 nor 3;6. The only reliable predictor was age, with the older preschoolers producing SO word order more often than the younger. The older children produced SO word order 99.5% of the time, whereas the younger did so about 85% of the time. These results are in line with the account of Hendriks et al. (2005), whose OT model of S-O word order acquisition predicts that an SO word order is optimal for the production of each of the four sentence types.

There were less children able to produce scorable utterances in the elicitation task in Experiment 1b than were able to produce scorable actions in Experiment 1a. Only five children remained in the younger group after two-thirds of those who participated were excluded from the production analysis. In addition to seeing whether a greater increase in animacy between subject and object would make a difference in children's production of S-O word order, the next experiment was carried out with twice as many children to counter the estimated 33% drop-out rate during the elicitation task.

## 5.3 Experiment 2b

### 5.3.1 Method

*Participants*

The same Dutch 2½-year-olds ($n = 32$) and 3½-year-olds ($n = 33$) who participated in the act-out task in Experiment 2a also participated in an elicitation task.

*Design and materials*

The same sentences with humans as animate nouns were tested as in Experiment 2a, and the same toys were used. As in Experiment 1b, no adjacent list items shared a toy or a sentence type, and no filler items were included.

*Procedure*

The same testing procedure was employed as in Experiment 1b.

*Scoring*

The same scoring procedure was employed as in Experiment 1b. An independent transcription of 10% of the participants ($n = 6$) showed 86% agreement with the main transcription. Inter-scorer agreement was also high (Cohen's $\kappa = .88$).

### 5.3.2 Results

*Criteria for inclusion*

There were 32 children excluded from the analysis because they were unable to contribute at least two scorable responses per sentence type. (Two of these children had been excluded from the analysis in 2a for the same reason). The removal of these participants decreased the number of 2½-year-olds by 22 (*n* = 10, *mean* = 2;9, *sd* = 1.8 mo.s). The number of 3½-year-olds decreased by 10 (*n* = 23, *mean* = 3;5, *sd* = 2.3 mo.s).

*Scorability*

The utterances produced by the remaining children were largely scorable. Of the 158 responses given by 2½-year-olds 121 were scorable and 37 were unscorable: 25 *Nontarget*, 4 *Insufficient*, and 8 *Unclear*. (2 received *No answer*.) Of the 364 responses given by 3½-year-olds, 332 were scorable, 32 were unscorable: 5 *Nontarget*, 4 *Insufficient*, and 23 *Unclear*. (4 received *No answer*.)

*Accuracy*

Mean accuracy for each group on the four sentence types based on participant means are shown in Table 5.3 and Figure 5.2. On average, the 2½-year-olds used *SO order* 81% (and *OS order* 19%). The 3½-year-olds used *SO order* 95% of the time (and *OS order* 5%). (Individual scores can be found in Table C.7 and Table C.8. in Appendix C.)

**Table 5.3 Mean proportion SO order produced: Experiment 2b**

| Sentence type | 2½yrs | *sd* | 3½yrs | *sd* |
|---|---|---|---|---|
| + animate + animate | .78 | *.37* | .90 | *.18* |
| + animate - animate | .87 | *.19* | .96 | *.12* |
| - animate + animate | .76 | *.28* | .98 | *.10* |
| - animate - animate | .83 | *.25* | .95 | *.13* |
| Total | .81 | *.19* | .95 | *.06* |

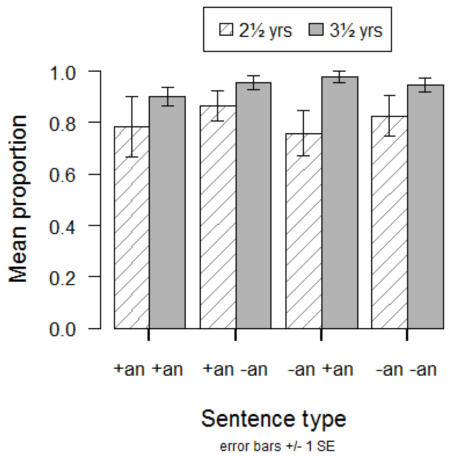


**Figure 5.2 Mean proportion SO order produced: Experiment 2b**

To determine whether S-O animacy affected accuracy in production in either age group, the binomial data were fit to a model with subject animacy, object animacy, and age group as fixed factors and participants and items as random factors. There was no significant three-way ($\chi^2(1) = 1.93$, $p > .1$) or two-way interaction ($\chi^2(3) = 4.36$, $p > .1$) between the fixed predictors. In the baseline model, there was no main effect of subject animacy ($\chi^2(1) = .75$, $p > .1$) or object animacy ($\chi^2(1) = 1.03$, $p > .1$). Age group was a significant predictor of accuracy ($\chi^2(1) = 9.05$, $p = .003$), with the older children producing SO word order more often than the younger children ($\beta = .83$; $z = 3.29$; $p < .001$). The inclusion of control factors such as gender, test verb, direction of action, and list did not significantly explain more variance in the data. Thus, the 3½-year-olds used SO order more often than the 2½-year-olds, and the animacy manipulation had no significant effect on either group.

### 5.3.3 Discussion

The results of Experiment 2b show that children's responses in the elicitation task did not vary as a function of the manipulation of subject and object animacy, neither at age 2;6 nor 3;6. The only reliable predictor was age, with the older preschoolers producing SO word order more often than the younger. The older children produced SO word order 95% of the time, whereas the younger did so 81% of the time. These results are in line with the results of Experiment 1b, which indicates that the greater difference in animacy between subject and object made no difference for the production of S-O word order by preschoolers.

## 5.4 General Discussion

The questions that Experiments 1b and 2b aim to answer is whether Dutch preschoolers' word order is adult-like and whether it is unaffected by the S-O animacy manipulation in the production task. The results of Experiments 1b and 2b are summed up as follows:

- Experiment 1b looked at the production of word order by 30 preschoolers (from Experiment 1a) using an elicitation task with sentences in which toy animals and vehicles interacted.
    - Age was a significant predictor, with the 3½-year-olds producing SO word order more often than the 2½-year-olds (99.5% vs. 85%).
    - No variation due to S-O animacy was found in the word order production of the 2½- or 3½-year-olds tested.

- Experiment 2b looked at the production of 65 preschoolers (from Experiment 2a) using an elicitation task with sentences in which toy humans and vehicles interacted.
    - Age was a significant predictor, with the 3½-year-olds producing SO word order more often than the 2½-year-olds (95% vs. 81%).
    - No variation due to S-O animacy was found in the word order production of the 2½- or 3½-year-olds tested.

Thus, both predictions were met: production was adult-like and animacy had no effect on S-O order in sentence production.

The mean scores on comprehension versus the mean scores on production in both experiments suggest that production exceeds comprehension in each age group. Word order was used 59% – 62% of the time by 2½-year-olds in comprehension and 81% – 85% of the time in production. Word order was used 83% – 92% of the time by 3½-year-olds in comprehension and 95% – 99% of the time in production. An analysis that compares scorability and accuracy between these comprehension and production tasks is presented in Chapter 8.

There were less children in the production analysis than in the comprehension analysis because of unscorable and missing responses in production. This is seen as a byproduct of an elicited production task, which allows participants freedom in how they respond. While C&M administered production items to children in as many as three sessions before getting scorable responses, the current paradigm paints a picture of the children's abilities that is more conservative.

To conclude, the results of Experiments 1 and 2 provide a first look at how Dutch children interpret and produce word order in the face of conflicting animacy cues. The experiments were designed to test only the effect of S-O animacy by carefully controlling for the effect of verb selectional restraints, and they provided evidence that (i) object animacy does influence sentence interpretation by 3½-year-olds and (ii) 2½- and 3½-year-old children can use SO order in their own speech. In order to determine whether this finding—that variation due to animacy is limited to comprehension—extends to other paradigms, the sentences from Experiment 1 were tested in two additional experiments. Comprehension tasks using cartoon animations are reported in Chapter 6, and their corresponding production tasks are reported in Chapter 7.

# 6 Comprehension: Picture selection & preferential looking with animations

Experiments 3a, 4a, 5a, & 6a

In the previous two chapters, two comprehension and two production experiments were presented in which Dutch preschoolers acted out sentences with toys or described actions they saw performed with toys. This chapter presents comprehension experiments in which Dutch- and English- preschoolers are tested with a picture selection task and a preferential looking task with cartoon animations as stimuli. Dutch- and English-speaking adults are additionally tested as controls. Comprehension experiments are reported in this chapter, one for each of the four populations tested. The accompanying sentence elicitation experiments carried out with the same populations and materials are presented in the next chapter. (An overview of all experiments is provided in Table 1.1 at the end of Chapter 1.)

## 6.1 Introduction

Picture selection and preferential looking have been used to test young children's comprehension of S-O word order, but these have not included animacy as a manipulation. Rather, sentences with two animals interacting are used to show that children could interpret S-O word order alone (Chan et al., 2010; Gertner et al., 2006; Hirsh-Pasek & Golinkoff, 1996a). Picture selection and preferential looking have an advantage over act-out tasks in that they restrict the possible interpretations of the sentences being tested. The experiments presented in this chapter test word order comprehension using the same sentences that were tested in Experiment 1, which used an animal-vehicle animacy distinction. In the two tasks used to test children, one of two animations is preferred via either point or gaze.

The first question that Experiments 3a - 6a aim to answer is whether a pattern of variable comprehension is found in a well-controlled picture selection and preferential looking task carried out with Dutch- and English-speaking preschoolers. Hendriks, de Hoop, and Lamers' (2005) model of early S-O word order predicts variable comprehension in both Dutch- and English-speaking preschoolers, whose grammar initially gives too much priority to an animacy constraint on meaning. I have predicted that variable comprehension will be evident in children of both languages at age 2;6, but will further persist in Dutch children until the age 3;6. By testing adult control groups, Experiments 3a and 5a test the predictions of the model of adult grammar proposed by de Hoop and Lamers (2006) and Hendriks et al. (2005), namely that adults of English and Dutch will rely on word order over animacy in sentence interpretation.

The second question to be answered is whether adults show processing effects due to S-O animacy. My model that implements de Hoop and Lamers' (2006) incremental constraint optimization predicts that sentences with inanimate subjects, especially S<O sentences, are at a disadvantage during

processing. Such effects, if found, would suggest that there is a lowly-ranked animacy constraint present in the grammar of adult speakers of Dutch and English.

## 6.2 Experiment 3a

### 6.2.1 Method

*Participants*

Native Dutch-speaking adults were tested (*n* = 41, 12 male, *mean* = 22 years). Participants were students or employees of the University of Groningen who volunteered to participate.

*Design and materials*

The 16 test items used were identical to those used in Experiment 1, which consisted of 8 *duwen* (*push*) and 8 *trekken* (*pull*) sentences with animals and vehicles (listed in Table A.2 in Appendix A). The only exception was that all instances of the word *trein* (*train*) were replaced by the word *bus* (*bus*). Each sentence type occurred four times each: [+an +an], [ +an -an], [-an +an], and [-an -an]. In addition to the test items, there were 16 filler items and 6 practice items (listed in Table A.5 in Appendix A), which included no test nouns or verbs. Thus, there were 38 sentences in total, which were pre-recorded with neutral prosody by a female voice. The experimental items (from either List A or B) were arranged in two different orders. No sentence type occurred twice in a row and no adjacent sentences shared a noun. List and order were distributed evenly across participant gender.

Colored animated movies depicting the action described by each test sentence served as the target animation and movies depicting the reversed action served as distractor animation. Animations of test items are shown in Figures B.7 – B.38 in Appendix B. The target and distractor animations appeared side by side on the screen. In half of the filler items, the action of the distractor animations depicted the reversed interpretation of the target sentence. In the other half of the fillers and in all of the practice items, the action of the distractor animations depicted a sentence with either a different entity as subject or a different entity as object than the target sentence (e.g. an animation of a balloon carrying a ball as the distractor for *The balloon is carrying the bottle*). Direction of action and target side was balanced across sentence types for both test and filler items.

*Apparatus*

The experimental items (pre-recorded sentences and animations) were sent to a Tobii T120 remote eye tracking monitor with resolution 1280 x 1024 by a computer running the software E-Prime® (Schneider, Eschman, & Zuccolotto, 2002) with E-Prime® Extensions for Tobii. The auditory stimuli were played on external desktop speakers. The accuracy, reaction time (RT), and gaze data (at a frame rate 60 Hz) was collected by E-Prime. Calibration was run with Tobii Studio Analysis Software™ on a second computer. This software also recorded the participant during testing via the eye tracking monitor’s embedded camera.

*Procedure*

Each participants’ eyes were first calibrated at five points on the screen. Then the picture selection task was administered either before or after the elicited production task to be described in Experiment 3b in Chapter 7. Instructions were given to press one of two marked keys on the left and right of the keyboard to indicate whether the animation on the left or the right of the screen matched the sentence. Participants then had three chances to practice. For each trial, the participant first saw a 1000 ms fixation cross in the center of the screen, which had to be fixated for 500 ms before the animation pair was shown. The pair of animations appeared on the screen for 300 ms before the sound file was played. The trial ended as soon as the participant pressed either key.

*Scoring*

For test items, if a participant chose an animation reflecting a SO interpretation, the item was scored as *SO interpretation*; if a participant chose an animation reflecting an OS interpretation, the item was scored as *OS interpretation*. If an item was not administered due to a technical error, it was scored as *Not given.*

### 6.2.2 Results

*Criteria for Inclusion*

All adults contributed at least two responses per sentence type. There was only 1 item *Not given* due to a technical error. In addition, it was clear from the low number of mistakes on the 16 filler items that they understood the task: 20 made no mistakes, 18 made 1 mistake, 3 made 2 mistakes.

*Accuracy*

There were 655 responses given on test items. Mean accuracy (*SO interpretations*) for the picture selection task on the four sentence types based on participant means are shown in Table 6.1. On average, adults selected animations that reflected *SO interpretations* 96.8% of the time (and *OS interpretations* 3.2%).

**Table 6.1 Mean proportion SO interpretations: Experiment 3a (Dutch adults)**

| Sentence type | Mean | *sd* |
|---|---|---|
| + animate + animate | .976 | *.08* |
| + animate - animate | .994 | *.04* |
| - animate + animate | .939 | *.12* |
| - animate - animate | .963 | *.09* |
| Total | .968 | *.04* |

To determine whether S-O animacy affected accuracy in comprehension, the binomial data were fit to a mixed effects model with subject and object animacy as fixed factors and participants and items as random factors. Since there was no significant two-way interaction ($\chi^2(1) = .46$, $p > .1$), main effects were checked for in the baseline model. There was no main effect of object animacy ($\chi^2(1) = 2.19$, $p > .1$), but there was a main effect of subject animacy ($\chi^2(1) = 4.65$, $p = .03$), with lower accuracy on sentences with an inanimate subject ($\beta = -.61$; $z = -2.08$; $p = .04$). The inclusion of control factors such as test verb, first task, target side, and list did not significantly explain more variance in the data. Thus, the Dutch adults gave SO interpretations to the sentences they heard about 95% of the time on average, but were less likely to do so when the subject was inanimate.

*Reaction time*

Items with OS interpretations ($n$ = 21) or extreme RTs ($n$ = 2) were removed from the RT analysis. Extreme RTs were defined as those outside 3 standard deviations of the participant's personal mean. Mean RT on the four sentence types based on participant means are shown in Table 6.2 and Figure 6.1. It took the Dutch adults 2140 ms on average to give an answer.

**Table 6.2 Mean RT for giving SO interpretation: Experiment 3a (Dutch adults)**

| Sentence type | RT(ms) | *sd* |
|---|---|---|
| + animate + animate | 2001 | *392* |
| + animate - animate | 1973 | *391* |
| - animate + animate | 2301 | *526* |
| - animate - animate | 2294 | *591* |
| Total | 2140 | *451* |

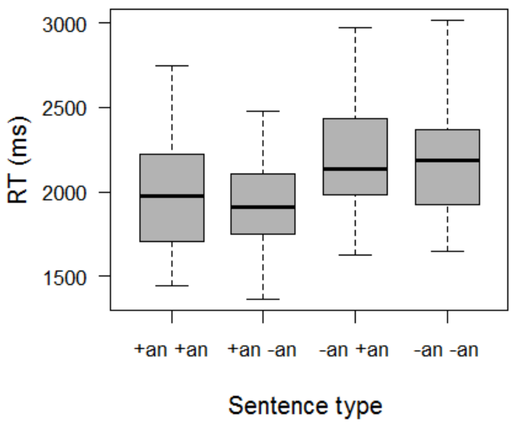


**Figure 6.1 Mean RT for giving SO interpretation: Experiment 3a (Dutch adults)**

To determine whether S-O animacy affected speed of response on the picture selection task, the log transformed RTs were fit to a model with subject and object animacy as fixed factors and participants and items as random factors. Since including an interaction was not justified ($\chi^2(1) = 0$, $p = 1$), the baseline model was checked for main effects. There was no main effect of object animacy ($\chi^2(1) = 0.10$, $p > .1$), but there was a main effect of subject animacy ($\chi^2(1) = 15.68$, $p < .001$). The adults had longer RTs when the subject was inanimate ($\beta = .07$; $t = 4.35$; $p < .001$). The inclusion of control factors such as test verb, first task, target side, and list did not significantly explain more variance in the data. Thus, the Dutch adults were faster to select the animation corresponding to a SO interpretation when the subject was animate than when the subject was inanimate.

*Eye movements between animations*

Items that remained in the RT analysis, but that had extreme track loss ($n = 3$) were removed from the analysis of eye movements. An item had extreme track loss if there was track loss of both eyes for more than one-third of the trial (from presentation to button press).

Areas of Interest (AOIs) in the visual stimuli were defined over *Target* animation, *Distractor* animation, and *Not on AOI*. Figure 6.2 shows the general pattern of looks to target and distractor over the course of a trial, collapsed over the four conditions, synchronized to the offset of the sentence subject (marked by 0 on the plot's x-axis). The gaze plot shows that Dutch adults looked to the target within 1000 ms of having heard the subject of the sentence.

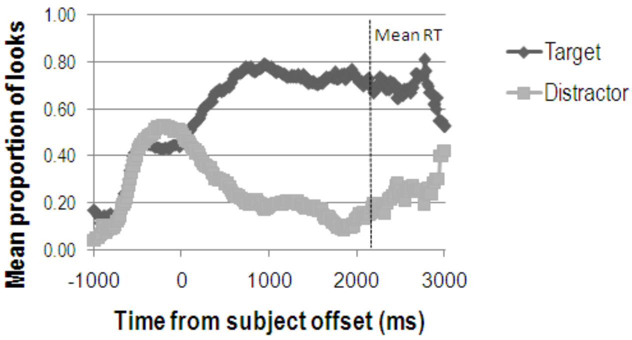


**Figure 6.2 Proportion of looks to target and distractor: Experiment 3a (Dutch adults)**

Four windows of time were defined for analysis: Time window 1 contains gaze data from the start of the trial to the offset of the sentence subject, with the subject duration about 600 ms. Time windows 2 – 4 are subsequent regions of 1000 ms following the offset of the sentence subject. For each participant and item, the proportion of target looks (versus distractor looks) during each of the four windows was calculated. Figure 6.3 plots the mean proportions of looks to target versus distractor in each of the time windows.

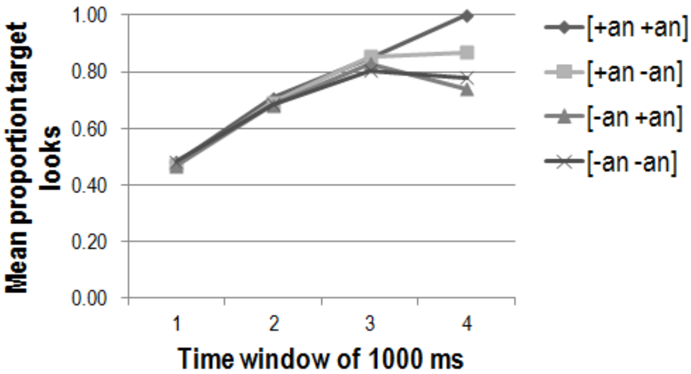


**Figure 6.3 Proportion of looks to target over four time windows: Experiment 3a (Dutch adults)**

To determine whether S-O animacy affected which AOI was fixated during picture selection, the empirical logit[17] transformed (Agresti, 2002: 87) mean looks to target from each time window were fit to a model with subject animacy, object animacy, and time window as fixed factors, and with participant and item as random factors. There was no significant three-way interaction between the fixed predictors ($\chi^2(1) = .18$, $p > .1$). There was a significant interaction of time window and subject animacy ($\chi^2(1) = 11.90$, $p < .001$) as well as a significant main effect of time window ($\chi^2(1) = 711.94$, $p < .001$). Thus, the adults looked increasingly towards the target animation as each time window progressed in general ($\beta = 1.98$; $t = 30.09$; $p < .001$), but to a significantly lesser degree when the subject was inanimate ($\beta = -.22$; $t = 3.45$; $p < .001$). The inclusion of control factors such as test verb, first task, target side, and list showed that target side significantly explained more variance in the data ($\chi^2(1) = 38.66$, $p < .001$), with participants more

[17] The empirical logit transformation is used in order to stay as close as possible to the values of a logit transformation, but avoiding the negative and positive infinity values for proportions of 0 and 1 respectively (Jaeger, 2008: 442)

likely to fixate on the target if it was on the left ($\beta$ = .48; $t$ = 8.04; $p$ < .001). The inclusion of an interaction of target side and time window proved also to be a significant improvement ($\chi^2(1) = 101.71$, $p$ < .001) indicating that the effect of target side decreased as the time window increased ($\beta$ = -.64; $t$ = 10.24; $p$ < .001).

*Eye movements within animations*

AOIs were further defined within target and distractor animations as either the agent of the action or the patient of the action. The difference between mean proportions of looks to agent and mean proportions of looks to the patient (or *agent advantage score*) within either the target or distractor animation, for each sentence type, based on participant means are listed in Table 6.3. A positive value indicates a preference for the agent, while a negative score indicates a preference for the patient. The mean agent advantage score for sentence types with no S-O contrast (S=O) are shown in Figure 6.4. For these sentences, the agent and patient were either both animate or both inanimate. The agent advantage score for sentence types with an S-O contrast (S≠O) are shown in Figure 6.5. For the [+an -an] sentences, the agent in the target animation was animate, and the agent in the distractor animation was inanimate; in contrast, for the [-an +an] sentences, the agent in the target animation was inanimate, and the agent in the distractor animation was animate.

**Table 6.3 Mean agent advantage score for target and distractor over trial: Experiment 3a (Dutch adults)**

| Sentence type | Agent advantage score **within target** | *sd* | Agent advantage score **within distractor** animation | *sd* |
|---|---|---|---|---|
| + animate + animate | .20 | *.19* | -.03 | *.25* |
| + animate - animate | .35 | *.19* | -.14 | *.26* |
| - animate + animate | -.06 | *.19* | .28 | *.24* |
| - animate - animate | .19 | *.20* | -.02 | *.20* |
| Total | .18 | *.12* | .03 | *.08* |

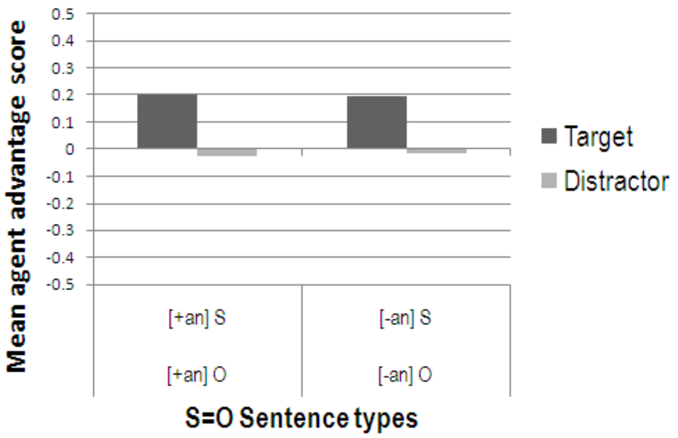


**Figure 6.4 Mean agent advantage score for S=O sentences: Experiment 3a (Dutch adults)**

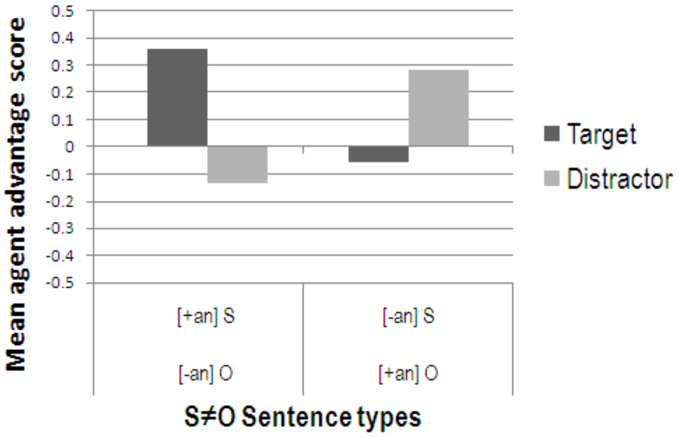


**Figure 6.5 Mean agent advantage score for S≠O sentences: Experiment 3a (Dutch adults)**

To determine whether agent animacy affected the degree to which the agent was fixated in either target or distractor animation during the trials with S=O sentences, the empirical logit transformed agent advantage scores for each participant and item were fit to a model with agent animacy, and animation (target vs. distractor) as fixed factors, and with participant and item as random factors. The empirical logit of the agent advantage scores was calculated by subtracting the empirical logit of looks to the patient from the empirical logit of looks to the agent during the trial. There was no significant two-way interaction between agent animacy and animation ($\chi^2(1) = .01$, $p > .1$), so main effects were checked for in the baseline model. There was a significant main effect of animation ($\chi^2(1) = 10.08$, $p = .001$), with the adults more likely to fixate the agent that was in the target animation versus the agent in the distractor animation ($\beta = .82$; $t = 3.18$; $p = .002$). Thus participants looked to agents in the target animations regardless of whether they were [+an +an] or [-an -an] trials.

The same analysis was carried out for trials with S≠O sentences. There was a significant interaction between agent animacy and animation ($\chi^2(1) = 54.29$, $p < .001$). Thus, the Dutch adults were more likely to fixate the agent that was within the target animation only in the [+an -an] sentences, in which the agent was animate ($\beta = 1.84$; $t = 7.53$; $p < .001$).

### 6.2.3 Discussion

Results of Experiment 3a show that Dutch adults' interpretations, RTs, and eye movements varied as a function of the manipulation of subject animacy. While they gave answers on the picture selection task that indicated a SO interpretation about 97% of the time on average, they did so more often when the subject was animate than when it was inanimate. Furthermore, on items in which they had SO readings, Dutch adults were faster to find the target when the subject was animate. This was reflected in both the speed of their button presses, as well as the degree to which they preferred to look to the target within the three seconds of having heard the sentence subject. The Dutch adults exhibited a pattern of scanning from left to right. They also looked to the agents in the target and distractor animations, a preference which was overridden when viewing [-an +an] animations.

No variable performance was expected in Dutch adults' interpretations, as the proposed ranking of a word order constraint over an animacy constraint for adult-speakers of Dutch predicts SO interpretations across the board. However, there was a significant difference between their SO answers on [+an] subject sentences (98.5%) and on [-an] subject sentences (95.1%). While on the whole, it is clear that Dutch adults interpreted sentences overwhelmingly as SO, subject animacy was playing a significant role in their offline sentence comprehension. Facilitation effects due to subject animacy *were* expected, and were reflected in the gaze and RT measures of sentences with animate subjects in Experiment 3a. We now turn to Dutch preschoolers to see if they also exhibit effects of S-O animacy in the comprehension tasks.

## 6.3 Experiment 4a

### 6.3.1 Method

*Participants*

Monolingual Dutch-speaking 2½-year-olds (*n* = 15, 5 male, range 2;5 – 3;2, *mean* = 2;9, *sd* = 2.8 mo.s) and 3½-year-olds (*n* = 17, 7 male, range 3;3 – 4;1, *mean* = 3;8, *sd* = 3.1 mo.s) participated in the study. Children were tested at the University of Groningen, and were selected from a database containing the contact information of parents in the Groningen area interested in participating in studies in the eye lab. Parents of children between 2;6 and 3;2 filled in the N-CDI III checklist, and parents of children aged 3;3 or older filled out the KINT checklist (see Section 4.2.1 for a description of the checklists).

*Design and materials*

The 16 experimental items (both animations and sentences) were identical to those in Experiment 3a, also with regard to correct side and direction of action. Easier practice items were used and only 4 fillers were included (listed in Table A.6 in Appendix A). No adjacent list items shared a noun or a sentence type.

*Apparatus*

The computer setup was the same as in Experiment 3a.

*Procedure*

The researcher and a native Dutch-speaking assistant invited a child and parent into a lab at the university. Before the comprehension tasks were begun, two pre-tests were carried out. The child was first asked to name pictures of each of the 19 nouns appearing in the animations. If the child did not know a word or used a different word, the assistant gave feedback and made sure to present the picture again so the child was familiar with it. The second pre-test involved a demonstration of the actions of pushing and pulling using a small toy pig and truck. Each had a string tied to them, intended to represent the rope depicted in the animations. The assistant first pushed the pig from behind and said "Look, this is 'pushing'. Now you try." The assistant indicated to the child to demonstrate *pushing* with the truck, while carefully avoiding any modeling of S-O word order. Next, the assistant pulled the truck and said "Look, this is 'pulling.'" As the child successfully pulled the truck, the assistant repeated "Yes, that's 'pulling', very good!" If a child had difficulty, the assistant used different strategies to guide the child until the child succeeded.

The child and parent then moved to a large reclining chair in front of the eye tracking monitor, where the child sat upright on the parent's lap. The parent wore dark glasses that kept the eye tracker from registering the parent's eyes and prevented any unconscious influencing of the child's answers. Once the child was positioned about 70 centimeters from the screen, the researcher ran a 5-point calibration.

The child was then tested in two blocks with a small break in between. During the first block the child was tested with a preferential looking task; during the second block the child was tested with a picture selection task. In each block, only one of the test verbs was tested. Whether *push* or *pull* was tested in the first block was balanced over list (A or B) and gender. During each block, the child was presented with trials (described below) in which the target and distractor animations appeared side-by-side on the screen as the pre-recorded sentences were played. During the a preferential looking task, the child was instructed to simply watch the video as their gaze data was collected. Each child saw 2 practice items (or warm-up items) followed by 8 test items with 2 interspersed filler items. During the picture selection task, the child was instructed to point to the animation that matched the sentence heard. Each child received 3 practice items followed by 8 test items with 2 interspersed filler items. The assistant gave instructive feedback during picture selection practice items if necessary, none of which contained distractors with a reversed interpretation of the target. The test sentence was repeated by the assistant if necessary during the pointing task. The entire session was repeated about one week later with the remaining items so that each child received both verbs with both tasks.

The timeline of a trial is shown in Figure 6.6. Each trial was preceded by an attention getting video together with an attractive sound for 2000 ms. This was followed by a gaze contingent cross in the center of the screen set to 100 ms. That is, the following slide would not appear until the child fixated on the cross for 100 ms. A baseline then appeared for 2500 ms, which presented the same visual stimuli as would appear during the test phase but without auditory input. The baseline was followed by a second gaze contingent fixation cross in the middle of the screen. The line that divided the animations remained on the screen during this re-fixation cross to avoid a distracting blinking effect. The re-fixation cross was red to increase its visibility on this line. Once this cross had been fixated for 100 ms, the test phase began. The animations were displayed again, this time together with a sound file that contained the test sentence, an exciting statement, and a repetition of the test sentence: *De auto duwt de koe. Wow! De auto duwt de koe.* There was a gap of 500 ms between each sentence. The first test sentence was identical to the second instance of the test sentence. The four interjections were balanced across sentence types: *Dat is leuk!, Kijk eens aan!, Hee!, Kijk!* (*Wow, Look at that, Hey, Look!*) In the preferential looking task, the sentence was played and the animations were visible for 7000 ms. In the picture selection task, the sentence was played and the animations were visible until the child made a decision by pointing gesture.

During the break the child and parent were able to leave the chair for five minutes. The child played with a toy and drank some juice. At the end of the break, the assistant briefly practiced pointing with the child. The assistant held up two of the pictures used during the naming pre-test and named one, which the child was asked to point to. This process was repeated at least three times. At the end of the second session the child was offered a small toy.

*Scoring*

The pointing responses were recorded by both the assistant via mouse click and by the researcher on a score sheet during testing. Discrepancies were rare and were resolved by consulting the session video.

A point to the target picture was scored as an *SO interpretation*, a point to the distractor was scored as an *OS interpretation*. If a child changed his or her mind, the final response was scored. Points clearly intended to signify both pictures (e.g. reinforced with a verbal “both” or “that one and that one”) were scored as *Both* and were treated as unscorable. Any items that received no response from the child were marked as *No answer*, and those not administered because the child was unwilling to continue (or due to a technical error) were marked as *Not given*.

### 6.3.1 Results

#### *6.3.1.1 Picture selection*

*Criteria for inclusion*

There were 2 children excluded from the analysis because they did not contribute at least two scorable responses per sentence type. The removal of these participants decreased the number of 2½-year-olds by 2

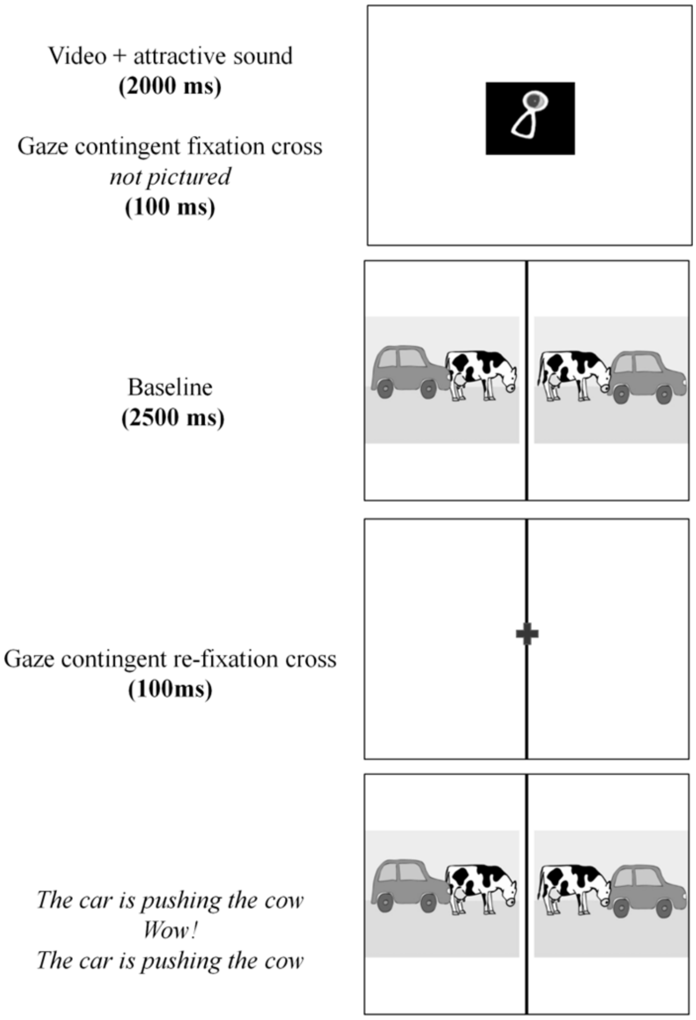


**Figure 6.6 Timeline of trial with animation pairs for preferential looking and picture-selection tasks**

(*n* = 13, *mean* = 2;9, *sd* = 3.0 mo.s). The number of 3½-year-olds did not change (*n* = 17, *mean* = 3;8, *sd* = 3.1 mo.s).

*Scorability*

The points by the children were largely scorable. Of the 205 responses given by the 2½-year-olds, 195 items were scorable and 10 were unscorable: 4 *Both* and 6 *Unclear*. (3 received *No answer*.) Of the 270 responses given by the 3½-year-olds, 266 items were scorable and 4 were unscorable: 3 *Both* and 1 *Unclear*. (1 received *No answer* and 1 was *Not given.*)

*Accuracy*

Mean accuracy for each group on the four sentence types based on participant means are shown in Table 6.4 and Figure 6.7. On average, the 2½-year-olds's responses reflected *SO interpretations* 54% of the time (and *OS interpretations* 46%), while the 3½-year-olds's responses reflected *SO interpretations* 70% of the time (and *OS interpretations* 30%). (Individual scores can be found in Table C.9 and Table C.10. in Appendix C.)

**Table 6.4 Mean proportion SO interpretations: Experiment 4a (Dutch preschoolers)**

| Sentence type | 2½yrs | *sd* | 3½yrs | *sd* |
|---|---|---|---|---|
| + animate + animate | .40 | *.25* | .73 | *.23* |
| + animate - animate | .72 | *.26* | .79 | *.25* |
| - animate + animate | .48 | *.27* | .54 | *.28* |
| - animate - animate | .54 | *.29* | .71 | *.27* |
| Total | .54 | *.10* | .70 | *.18* |

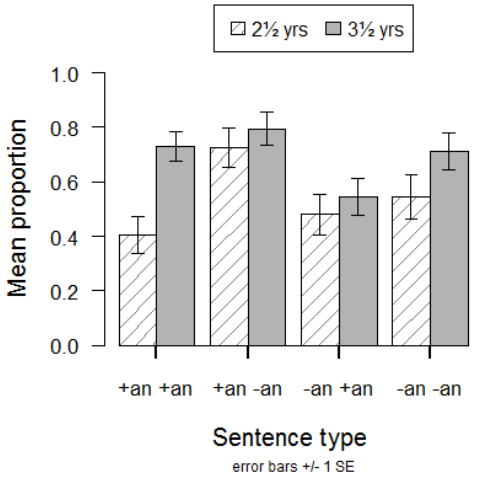


**Figure 6.7 Mean proportion SO interpretations: Experiment 4a (Dutch preschoolers)**

To determine whether S-O animacy affected accuracy in comprehension in either age group, the binomial data were fit to a model with subject animacy, object animacy, and age group as fixed factors, and participants and items as random factors. There were no significant three-way ($\chi^2(1) = 2.37$, $p > .1$) or two-way ($\chi^2(3) = 2.68$, $p > .1$) interactions between the fixed predictors. Since including interactions was not justified, the baseline model was checked for main effects. There were three distinct main effects that were significant predictors of accuracy. There was a main effect of age group ($\chi^2(1) = 8.20$, $p = .004$), with the older children more likely to choose SO animations than the younger children ($\beta = .37$; $z = 3.06$; $p = .002$); a main effect of subject animacy ($\chi^2(1) = 5.67$, $p = .02$), with all children more likely to choose SO animations when the subject was animate ($\beta = .24$; $z = 2.39$; $p = .02$); and a main effect of object animacy ($\chi^2(1) = 12.58$, $p < .001$), with all children less likely to choose SO animations when the object was animate ($\beta = -.36$; $z = -3.54$; $p < .001$). The inclusion of control factors such as gender, test verb, target side, list and vocabulary score did not significantly explain more variance in the data. Thus, both 3½-year-olds and 2½-year-olds gave more accurate answers when the subject was animate as well as when the object was inanimate.

#### *6.3.1.2 Preferential looking*

*Criteria for inclusion*

All 32 children who participated in the experiment were included in the preferential looking analysis because very few trials had extreme track loss. An item had extreme track loss if there was track loss of both eyes for more than one-third of the four-second region during the trial that would be analyzed. Test items with extreme track loss ($n = 70$) were removed from the analysis of eye movements. Each child contributed at least two validly tracked items per sentence type except for 5, who contributed at least two valid items on at least three sentence types.

*Eye movements between animations*

AOIs in the visual stimuli were defined over *Target* animation, *Distractor* animation, and *Not on AOI*. Eye movements during the baseline were first inspected to make sure that there was no preference for one type of picture over the other by the children that might affect how any effect of linguistic input can be interpreted. Table 6.5 shows the mean proportion of looks to target (versus distractor) during the 2500 ms baseline for each age group and sentence type based on participant means. Preference for the target during the baseline ranged between 43% and 54%.

**Table 6.5 Mean proportion target looks during baseline: Experiment 4a (Dutch preschoolers)**

| Sentence type | 2½yrs | *sd* | 3½yrs | *sd* |
|---|---|---|---|---|
| + animate + animate | .47 | *.21* | .42 | *.18* |
| + animate - animate | .43 | *.17* | .51 | *.17* |
| - animate + animate | .54 | *.18* | .48 | *.17* |
| - animate - animate | .53 | *.23* | .53 | *.18* |
| Total | .49 | *.09* | .50 | *.06* |

To determine whether S-O animacy depicted in the target animation affected whether it was fixated during the baseline for each age group, the empirical logit transformed mean looks to target were fit to a model with subject animacy, object animacy, and age group as fixed factors, and with participant and item as random factors. There were no significant three-way ($\chi^2(1) = .06$, $p > .1$) or two-way ($\chi^2(3) = 6.02$, $p > .1$) interactions between the fixed predictors. Furthermore, there was no main effect of subject animacy ($\chi^2(1) = .89$, $p > .1$), object animacy ($\chi^2(1) =.08$, $p > .1$), or age group ($\chi^2(1) = 1.78$, $p > .1$). Thus, neither the 2½-year-olds nor the 3½-year-olds showed a pre-existing preference for either target or distractor in any of the four sentence types. The inclusion of control factors such as gender, test verb, target side, and list showed that target side significantly explains more variance in the data ($\chi^2(1) = 42.24$, $p < .001$), with children more likely to fixate on the target if it was on the left ($\beta = .88$; $t = 8.16$; $p < .001$).

Figure 6.8 shows the general pattern of looks to target and distractor over the course of a trial, synchronized to the offset of the subject. The gaze plot shows that, in general, the overall proportions of looks to the target by Dutch preschoolers did not reach above .60.

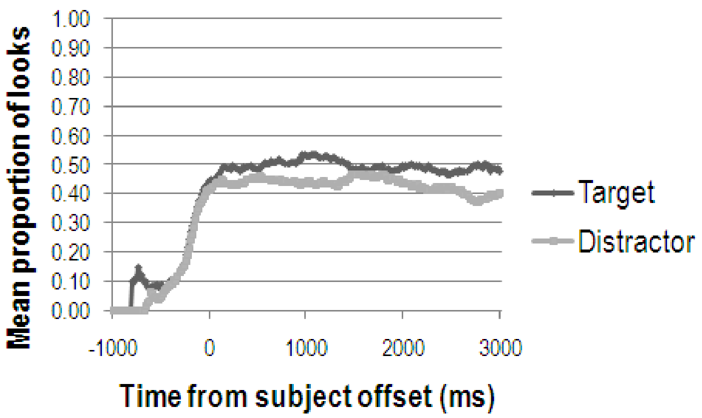


**Figure 6.8 Looks to target or distractor during preferential looking task by Dutch children**

Four windows of time were defined for analysis: Time window 1 contains gaze data from the start of the trial to the offset of the sentence subject, with the subject duration about 600 ms. Time windows 2 – 4 are subsequent regions of 1000 ms following the offset of the sentence subject. For each participant and item, the proportion of target looks (versus distractor looks) during each of the four windows was calculated. Figure 6.9 and Figure 6.10 plot the mean proportions of looks to target versus distractor in each of the time windows for each age group.

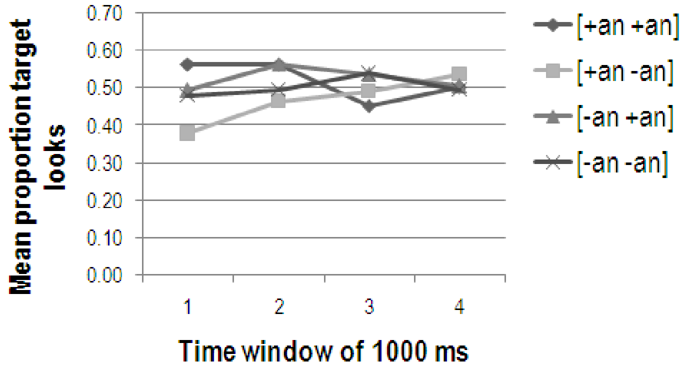


**Figure 6.9 Proportion of looks to target over four time windows: Experiment 4a (Dutch 2½ yrs)**

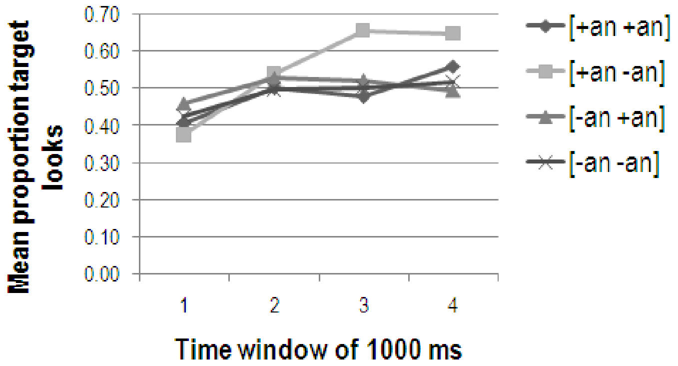


**Figure 6.10 Proportion of looks to target over four time windows: Experiment 4a (Dutch 3½ yrs)**

To determine whether S-O animacy affected which AOI was fixated during preferential looking of the 2½-year-olds, the empirical logit transformed mean looks to target were fit to a model with subject animacy, object animacy, and time window as fixed factors, and with participant and item as random factors. There were no three-way ($\chi^2(1) = .07$, $p > .1$) or two-way ($\chi^2(1) = 1.10$, $p > .1$) interactions between the fixed predictors. In the baseline model there was only a significant effect of time window ($\chi^2(1) = 19.58$, $p < .001$) indicating that the younger children looked increasingly towards the target animation as each time window progressed in general ($\beta = .34$; $t = 4.45$; $p < .001$). The inclusion of control factors such as gender, test verb, target side, and list showed that target side significantly explains more variance in the data ($\chi^2(1) = 21.36$, $p < .001$), with children more likely to fixate on the target if it was on the left ($\beta = .68$; $t = 5.05$; $p < .001$). Target side appears not to interact with time window ($\chi^2(1) = .10$; $p > .10$) as it had for Dutch adults.

The same analysis was carried out for the 3½-year-olds. There was no three-way ($\chi^2(1) = .97$, $p > .1$) interaction between the fixed predictors. There was a two-way interaction between time window and subject animacy ($\chi^2(1) = 4.39$, $p = .04$). Together with the main effect of time window ($\chi^2(1) = 27.61$, $p < .001$), this indicates that the older children looked increasingly towards the target animation as each time window progressed in general ($\beta = .36$; $t = 5.29$; $p < .001$), an effect that was intensified when the subject was animate ($\beta = .14$; $t = 2.10$; $p = .04$). Thus, the Dutch 3½-year-olds' looking reflected a preference for animate subjects, whereas the 2½-year-olds showed no such preference. The inclusion of control factors in the model of the older children, such as gender, test verb, target side, and list, did not significantly explain more variance in the data.

*Eye movements within animations*

AOIs were further defined within target and distractor animations as either the agent or the patient of the action. The agent advantage score within either the target or distractor animation, for each sentence type, based on participant means are listed for each age group in Table 6.6 and Table 6.7. The mean agent advantage scores for the younger group is shown in Figure 6.11 and for the older group in Figure 6.12 for sentence types with no S-O contrast (S=O). The mean agent advantage scores collapsed over age groups are shown in Figure 6.13 for sentence types with an S-O contrast (S≠O).

**Table 6.6 Mean agent advantage score for target and distractor over trial: Experiment 4a (Dutch 2½ yrs)**

| Sentence type | Agent advantage score **within target** | *sd* | Agent advantage score **within distractor** | *sd* |
|---|---|---|---|---|
| + animate + animate | .05 | *.31* | .06 | *.37* |
| + animate - animate | .17 | *.24* | -.09 | *.33* |
| - animate + animate | -.11 | *.42* | .25 | *.28* |
| - animate - animate | .17 | *.21* | .04 | *.20* |
| Total | .09 | *.15* | .06 | *.13* |

**Table 6.7 Mean agent advantage score for target and distractor over trial: Experiment 4a (Dutch 3½ yrs)**

| Sentence type | Agent advantage score **within target** | *sd* | Agent advantage score **within distractor** | *sd* |
|---|---|---|---|---|
| + animate + animate | .25 | *.40* | -.16 | *.34* |
| + animate - animate | .22 | *.35* | -.24 | *.32* |
| - animate + animate | -.03 | *.35* | .20 | *.45* |
| - animate - animate | .08 | *.38* | .16 | *.36* |
| Total | .10 | *.21* | -.02 | *.17* |

To determine whether agent animacy affected the degree to which the agent was fixated in either target or distractor animation during the trials with S=O sentences by either age group, the empirical logit transformed agent advantage scores for each participant and item were fit to a model with agent animacy, animation (target vs. distractor), and age group as fixed factors, and with participant and item as random factors. There was a significant three-way interaction between the fixed predictors ($\chi^2(1) = 4.30$; $p = .04$). The interaction picks up on a difference between the age groups ($\beta = .62$; $t = 2.06$; $p = .04$): the younger children in fact showed the greatest preference for agent over patient in the target animation in the [-an

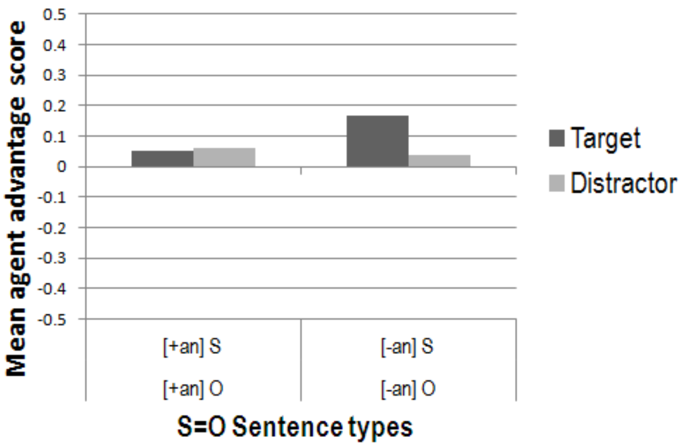


**Figure 6.11 Mean agent advantage score for S=O sentences: Experiment 3a (Dutch 2½ yrs)**

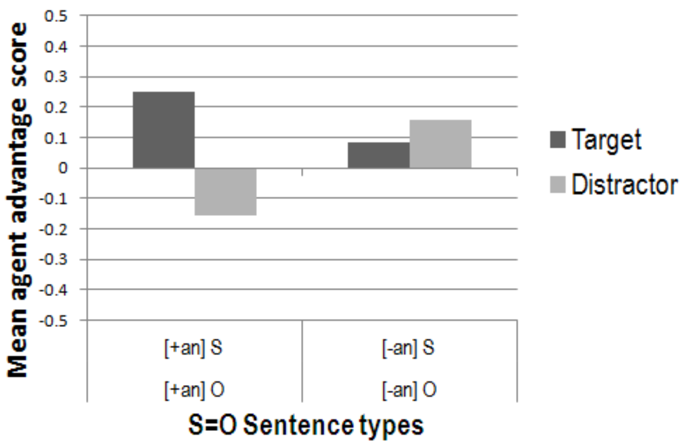


**Figure 6.12 Mean agent advantage score for S=O sentences: Experiment 3a (Dutch 3½ yrs )**

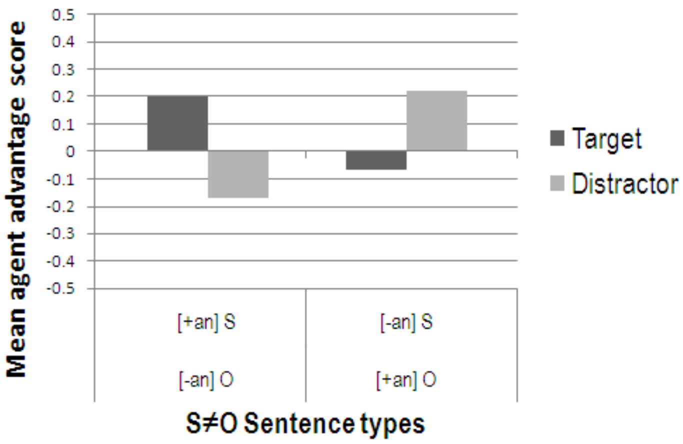


**Figure 6.13 Mean agent advantage score for S≠O sentences: Experiment 3a (Dutch preschoolers)**

-an] sentences, and the older children showed the greatest preference for agent over patient in the target animation in the [+an +an] animations. Thus, the Dutch children more or less preferred to look to the agents in the S=O targets and distractor animations, with 2½-year-olds favoring vehicle agents in the target animations and 3½-year-olds favoring animal agents in the target animations.

The same analysis was carried out for trials with S≠O sentences. There was no three-way interaction ($\chi^2(1) = .38$, $p > .1$), but there was a significant two-way interaction between agent animacy and animation ($\chi^2(1) = 29.18$; $p < .001$). Thus, the Dutch children were more likely to fixate the agent that was within the target animation only in the [+an -an] sentences, in which the agent was animate ($\beta = 1.39$; $t = 5.32$; $p < .001$).

### 6.3.2 Discussion

Results of the picture selection task in Experiment 4a show that Dutch preschoolers' interpretations varied as a function of the manipulation of both subject animacy and object animacy. The result of these two main effects was good comprehension of word order on the S>O sentences in both age groups (72% by younger children and 79% by older children) and poor use of word order on the S<O sentences (48% for younger children and 54% for older children). As in the act-out tasks, the older children's responses on the pointing task tended to reflect SO interpretations more often than those of the younger children. Thus, Dutch preschoolers showed variable comprehension on the picture selection task in precisely the way predicted by the OT model. Furthermore, this variable performance was found in both 2½-year-olds and 3½-year-olds, meeting the cross-linguistic prediction that Dutch children would fail to rely on word order by the age of 3;6.

Results of the preferential looking task in Experiment 4a show that the older Dutch preschoolers' gaze varied as a function of the manipulation of subject animacy. By the second and third second after subject offset, the 3½-year-olds showed a greater increase in looks to the target in the sentences with subjects that were animals than in the sentences with subjects that were vehicles. We can be certain that this was not due to a pre-existing preference for pictures in which animals are agents since the 2.5 second baseline revealed equal looking to target and distractor across sentence types. In fact, the baseline looking behavior was likely a scanning from left to right, evidenced by the significant effect of target side. The 2½-year-olds, in contrast to their older counterparts, did not exhibit gaze patterns that varied as a function of the animacy manipulation.

The analysis of within-animation looks showed a pattern that paralleled that of the Dutch adults. Remember that the Dutch adults showed a preference for the agent in the target animations when there was no contrast in animacy between the entities in the animations. The Dutch preschoolers, while having slightly different preferences for agents in target animations based on the different S=O sentence types, also generally looked to the agent in the target and distractor pictures when there was no contrast in animacy between the entities in the animations. When it came to the animations in which there was a contrast in animacy between the entities, both Dutch adults and preschoolers showed a preference for the agent in the target picture only in the [+an -an] sentences. The pattern of looks during S=O and S≠O sentences taken together show that there is an overall preference to look at the agent of the action during this type of comprehension task, a preference that is overridden when the agent of the action is inanimate and the patient of the action is animate. This within-animation gaze pattern will be discussed further at the end of the chapter. We turn now to the comprehension experiments run with English-speaking adults and preschoolers.

## 6.4 Experiment 5a

### 6.4.1 Method

*Participants*

Native English-speaking adults were tested (*n* = 31, 13 male, *mean* = 20 years). Participants were psychology students of the University of Pennsylvania who received course credit for their participation.

*Design and materials*

The same sentences, animations, and design were used as in Experiment 3a. English versions of the sentences were recorded with a neutral prosody by a female voice.

*Apparatus*

The equipment set-up was the same as in Experiment 3a except for the following differences: The eye tracking monitor had a resolution of 1024 x 768; the calibration procedure and the experiment were run from the same rather than separate computers; and no video recording was made of the participant.

*Procedure*

The same testing and scoring procedure was employed as in Experiment 3a.

### 6.4.2 Results

*Criteria for inclusion*

All adults contributed at least two responses per sentence type. There was only 1 item *Not given* due to a technical error. It was clear from the fairly low number of mistakes on the 16 filler items that they understood the task: 21 made no mistakes, 9 made 1 mistake,1 made 2 mistakes.

*Accuracy*

There were 495 responses given on test items. Mean accuracy on the picture selection task for the four sentence types based on participant means are shown in Table 6.8. On average, adults selected animations that reflected *SO interpretations* 97.4% of the time (and *OS interpretations* 2.6%).

**Table 6.8 Mean proportion SO interpretations: Experiment 5a (English-speaking adults)**

| Sentence type | Mean | *sd* |
|---|---|---|
| + animate + animate | .992 | *.04* |
| + animate - animate | 1.00 | - |
| - animate + animate | .968 | *.09* |
| - animate - animate | .935 | *.13* |
| Total | .974 | *.04* |

To determine whether S-O animacy affected accuracy in comprehension, the binomial data were fit to a model with subject and object animacy as fixed factors and participants and items as random factors. The fully complex model was not fit due to collinearity between the two-way interaction of the fixed predictors and object animacy, therefore only the baseline model was created. There was no main effect of object animacy ($\chi^2(1) = .07$, $p > .1$), but subject animacy made a significant contribution ($\chi^2(1) = 5.70$, $p = .01$), with a coefficient for subject animacy suggesting that there were fewer SO responses on sentences with an inanimate subject ($\beta$ = -1.59; $z$ = -2.6; $p$ = .11). The inclusion of control factors such as test verb, first task, target side, and list did not significantly explain more variance in the data. Thus, the English-speaking adults gave SO interpretations to the sentences they heard about 97% of the time on average, but were less likely to do so when the subject was inanimate.

*Reaction time*

Items with OS interpretations ($n$ = 13) or extreme RTs ($n$ = 4) were removed from the RT analysis. Mean RT's on the four sentence types based on participant means are shown in Table 6.9 and Figure 6.14. It took the English-speaking adults 2216 ms on average to give an answer.

**Table 6.9 Mean RT for giving SO interpretation: Experiment 5a (English-speaking adults)**

| Sentence type | RT(ms) | *sd* |
|---|---|---|
| + animate + animate | 2157 | *468.5* |
| + animate - animate | 2096 | *334.4* |
| - animate + animate | 2300 | *400.4* |
| - animate - animate | 2311 | *383.1* |
| Total | 2216 | *363.7* |

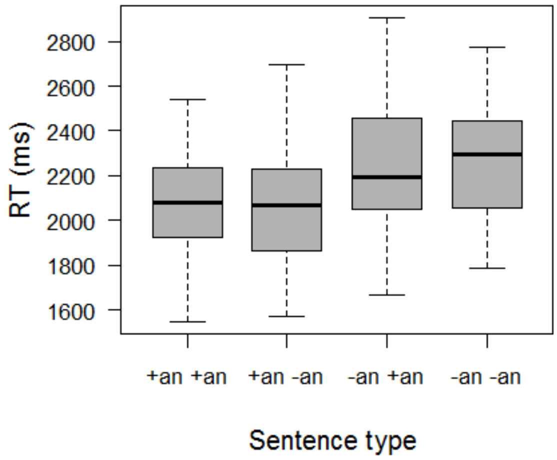


**Figure 6.14 Mean RT for giving SO interpretation: Experiment 5a (English-speaking adults)**

To determine whether S-O animacy affected speed of response on the comprehension task, the log transformed RTs were fit to a model with subject and object animacy as fixed factors and participants and items as random factors. Since there was no significant two-way interaction ($\chi^2(1)$ = .68, $p$ > .1), main effects were checked for in the baseline model. There was no main effect of object animacy ($\chi^2(1)$ = 0, $p$ = 1), but there was a main effect of subject animacy ($\chi^2(1)$ = 13.58, $p$ < .001). The adults had longer RTs when the subject was inanimate ($\beta$ = .04; $t$ = 3.92; $p$ < .001). The inclusion of control factors such as test verb, first task, target side, and list did not significantly explain more variance in the data. Thus, the English-speaking adults were faster to select the animation corresponding to a SO interpretation when the subject was animate than when the subject was inanimate.

*Eye movements between animations*

Items that remained in the RT analysis, but that had extreme track loss ($n = 7$) were removed from the analysis of eye movements. In addition, all items from one adult who had extreme track loss on most items was removed from the analysis, leaving 30 adults.

AOIs in the visual stimuli were defined over *Target* animation, *Distractor* animation, and *Not on AOI*. English-speaking adults showed the same general pattern of looking to the target shortly after the sentence started as the Dutch-speaking adults did. Four windows of time were defined for analysis: Time window 1 contains gaze data from the start of the trial to the offset of the sentence subject, with the subject duration about 500 ms. Time windows 2 – 4 are subsequent regions of 1000 ms following the offset of the sentence subject. For each participant and item, the proportion of target looks (versus distractor looks) during each of the four windows was calculated. Figure 6.15 plots the mean proportions of looks to target versus distractor in each of the time windows.

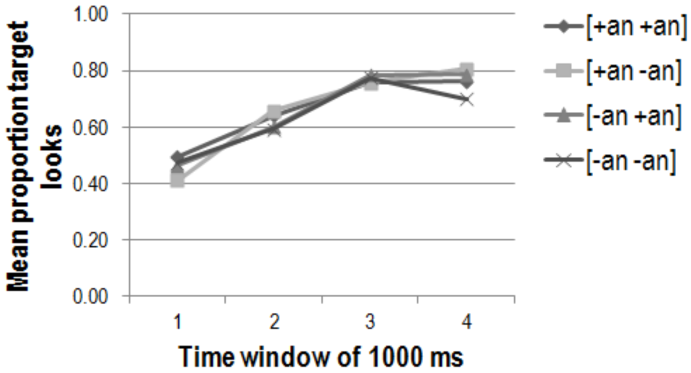


**Figure 6.14 Proportion of looks to target over four time windows: Experiment 5a (English-speaking adults)**

To determine whether S-O animacy affected which AOI was fixated during picture selection, the empirical logit transformed mean looks to target from each time window were fit to a model with subject animacy, object animacy, and time window as fixed factors, and with participant and item as random factors. There were no significant three-way ($\chi^2(1) = 2.68$, $p > .1$) or two-way ($\chi^2(3) = 0.70$, $p > .1$) interactions between the fixed predictors. In the baseline model, there was no main effect of subject animacy ($\chi^2(1) = .14$, $p > .1$) or object animacy ($\chi^2(1) = 0$, $p > .1$), but there was a main effect of time window ($\chi^2(1) = 482.08$, $p < .001$). Thus, the adults looked increasingly towards the target animation as each time window progressed in general ($\beta = 1.98$; $t = 24.08$; $p < .001$). Thus, the gaze of English-speaking adults seemed not to be affected by the animacy manipulation. The inclusion of control factors such as test verb, first task, target side, and list showed that target side significantly explained more variance in the data ($\chi^2(1) = 18.048$, $p < .001$), with participants more likely to fixate on the target if it was on the left ($\beta = .31$; $t = 4.48$; $p < .001$). The inclusion of an interaction of target side and time window proved also to be a significant improvement ($\chi^2(1) = 84.43$, $p < .001$) indicating that the effect of target side decreased as the time (window) increased ($\beta = -.70.$; $t = 9.34$; $p < .001$).

*Eye movements within animations*

AOIs were further defined within target and distractor animations as either the agent or patient of the action. The agent advantage score within either the target or distractor animation, for each sentence type, based on participant means are listed in Table 6.10. The mean agent advantage score for sentence types with no S-O contrast (S=O) are shown in Figure 6.16. and for sentence types with an S-O contrast (S≠O) in Figure 6.17.

**Table 6.10 Mean agent advantage score for target and distractor over trial: Experiment 5a (English-speaking adults)**

| Sentence type | Agent advantage score **within target** | *sd* | Agent advantage score **within distractor** | *sd* |
|---|---|---|---|---|
| + animate + animate | .21 | *.22* | .06 | *.32* |
| + animate - animate | .35 | *.21* | -.27 | *.21* |
| - animate + animate | -.04 | *.32* | .23 | *.25* |
| - animate - animate | .20 | *.24* | -.05 | *.38* |
| Total | .18 | *.11* | .01 | *.23* |

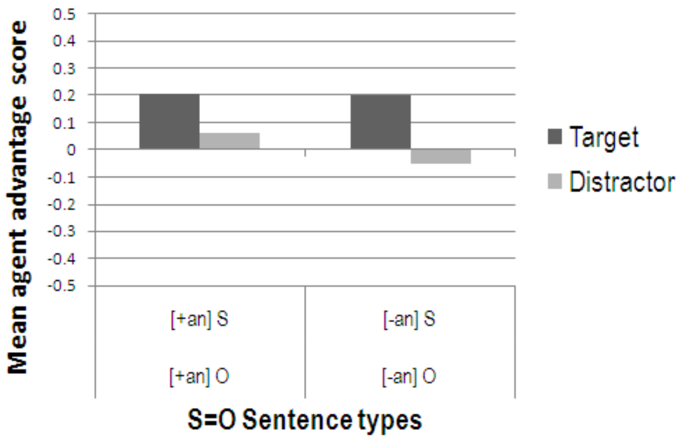


**Figure 6.16 Proportion of looks to agent and patient for S=O sentences: Experiment 5a (English-speaking adults)**

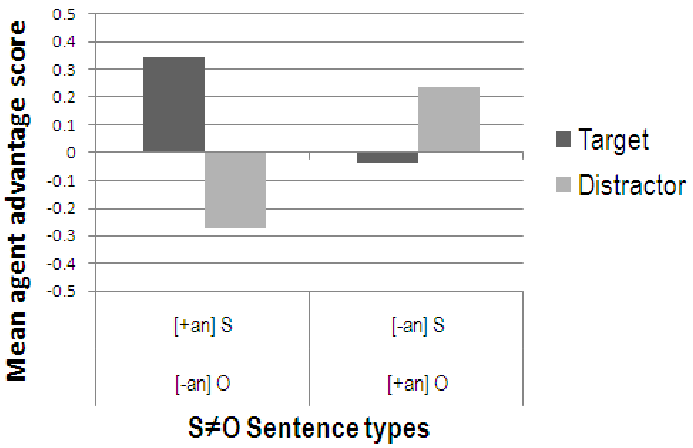


**Figure 6.17 Proportion of looks to agent and patient for S≠O sentences: Experiment 5a (English-speaking adults)**

To determine whether agent animacy affected the degree to which the agent was fixated in either target or distractor animation during the trials with S=O sentences, the empirical logit transformed agent advantage scores for each participant and item were fit to a model with agent animacy and animation (target vs. distractor) as fixed factors, and with participant and item as random factors. There was no significant two-way interaction between agent animacy and animation ($\chi^2(1) = .79$, $p > .1$), so main effects were checked for in the baseline model. There was a significant main effect of animation ($\chi^2(1) = 10.16$, $p = .001$), with the adults more likely to fixate the agent that was in the target animation versus the agent in the distractor animation ($\beta = .81$; $t = 3.20$; $p = .001$). Thus participants looked to agents in the target animations regardless of whether they were [+an +an] or [-an -an] trials.

The same analysis was carried out for trials with S≠O sentences. There was a significant interaction between agent animacy and animation (target vs. distractor) ($\chi^2(1) = 63.06$, $p < .001$). Thus, the English-speaking adults were more likely to fixate the agent that was within the target animation only in the [+an -an] sentences, in which the agent was animate ($\beta = .92$; $t = 7.53$; $p < .001$).

### 6.4.3 Discussion

Results of Experiment 5a show that English-speaking adults' interpretations and RTs—but not their eye movements—varied as a function of the manipulation of subject animacy. Like Dutch adults, the English adults gave answers on the picture selection task that indicated a SO interpretation about 97% of the time on average, and they did so more often when the subject was animate than when it was inanimate. Furthermore, on items in which they gave SO responses, English-speaking adults were faster to find the target when the subject was animate. This was reflected in the speed of their button presses only. Their between-animation eye movements exhibited a pattern of scanning from left to right, but revealed no increased preference to look to the target when the subject was animate. Their within-animation eye movements showed that they had an overall preference to look at agents, a preference that was overridden when viewing [-an +an] animations.

Like Dutch adults, English-speaking adults exhibited variable performance due to the animacy manipulation in the way they interpreted the meaning of S-O word order. This variation was found offline, which was not expected in Dutch or English since both languages are proposed to rank word order more highly than animacy in their grammars. English-speaking adults did interpret sentences they heard overwhelmingly as SO, but subject animacy played a significant role in their offline sentence comprehension. Furthermore, English-speaking adults exhibited facilitation effects expected during the online processing of sentences with animate subjects, found both in their RTs and within-animation gaze patterns. We now turn to the comprehension results from English-speaking preschoolers.

## 6.5 Experiment 6a

### 6.5.1 Method

*Participants*

Monolingual English-speaking 2½-year-olds ($n = 19$, 7 male, age range 2;3 – 3;2, *mean* = 2;9, *sd* = 3.7 mo.s) and 3½-year-olds ($n = 12$, 9 male, age range 3;4 – 3;11, *mean* = 3;9, *sd* = 2.0 mo.s) participated in the study. The children attended day cares in Philadelphia that cooperated with the University of Pennsylvania's Language Development and Language Processing Lab. No information about the vocabulary of the children was collected.

*Design and materials*

The same sentences (in English), animations, and design were used as in Experiment 4a.

*Apparatus*

The equipment setup was the same as in Experiment 5a.

*Procedure*

The same testing and scoring procedure was employed as in Experiment 4a except for the following differences. Since testing took place at day cares, parents were not present. Rather children were tested in a quiet room with the researcher and an assistant. One minor change is that a toy horse and truck were used in the demonstration of *push* and *pull*, rather than a toy pig and a truck.

### 6.5.2 Results

#### *6.5.2.1 Picture selection*

*Criteria for inclusion*

Three children were tested in only one session because they were not present when the researchers returned for their second session; however they contributed enough scorable responses to be included in the analysis. There was one child (aged 2;4) excluded from the analysis because she did not contribute at least two scorable responses per sentence type. Her removal decreased the number of 2½-year-olds by 1 (*n* = 18, *mean* = 2;9, *sd* = 3.6 mo.s). The number of 3½-year-olds did not change (*n* = 12, *mean* = 3;9, *sd* = 2.0 mo.s).

*Scorability*

The points by the children were largely scorable. Of the 271 responses given by the 2½-year-olds, 269 items were scorable and 2 were unscorable (*Unclear)*. (1 received *No answer* and 10 were *Not given.*) Of the 171 responses given by the 3½-year-olds, all were scorable. (2 received *No answer* and 3 were *Not given.*)

*Accuracy*

Mean accuracy for each group on the four sentence types based on participant means are shown in Table 6.11 and Figure 6.18. On average, the 2½-year-olds's responses reflected *SO interpretations* 60% of the time (and 40% for *OS interpretations*), while the 3½-year-olds's responses reflected *SO interpretations* 80% of the time (and *OS interpretations* 20%). (Individual scores can be found in Table C.11 and Table C.12. in Appendix C.)

**Table 6.11 Mean proportion SO interpretations: Experiment 6a (English-speaking preschoolers)**

| Sentence type | 2½yrs | *sd* | 3½yrs | *sd* |
|---|---|---|---|---|
| + animate + animate | .63 | *.28* | .77 | *.29* |
| + animate - animate | .65 | *.24* | .87 | *.14* |
| - animate + animate | .61 | *.20* | .81 | *.22* |
| - animate - animate | .52 | *.24* | .77 | *.20* |
| Total | .60 | *.15* | .80 | *.13* |

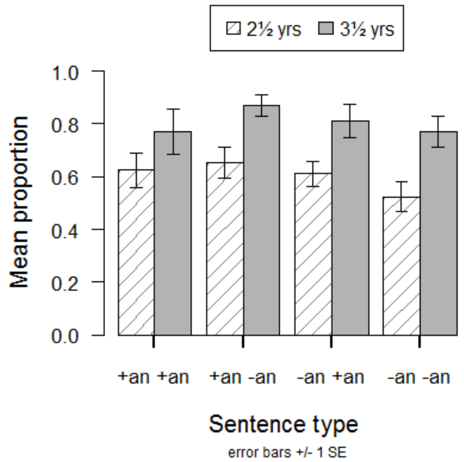


**Figure 6.18 Mean proportion SO interpretations: Experiment 6a (English-speaking preschoolers)**

To determine whether S-O animacy affected accuracy in comprehension in either age group, the binomial data were fit to a model with subject animacy, object animacy, and age group as fixed factors, and participants and items as random factors. There was no significant three-way ($\chi^2(1) = 0.44$, $p > .1$) or two-way interaction ($\chi^2(3) = 3.24$, $p > .1$) between the fixed predictors. In the baseline, there was no effect of subject animacy ($\chi^2(1) = .87$, $p > .1$) or object animacy ($\chi^2(1) = .11$, $p > .1$). There was a main effect of age group ($\chi^2(1) = 11.43$, $p < .001$), with the older children more likely to select SO animations than the younger children ($\beta = .52$; $z = 3.65$; $p < .001$). The inclusion of control factors such as gender, test verb, target side, and list showed that target side significantly explains more variance in the data ($\chi^2(1) = 3.73$, $p = .05$) with children more likely to answer correctly when the target was on the right ($\beta = .24$; $z = 2.00$; $p = .05$).

### *6.5.2.2 Preferential looking*

*Criteria for inclusion*

Of the 31 children who participated in the experiment, 1 child was excluded from the preferential looking analysis because of too much track loss. The number of 2½-year-olds remained the same ($n$ = 19, *mean* = 2;9, *sd* = 3.7 mo.s) and the 3½-year-olds decreased by 1 ($n$ = 11, *mean* = 3;8, *sd* = 2.0 mo.s). Items with extreme track loss ($n$ = 38) were additionally removed from the analysis of eye movements, but each remaining child contributed at least two validly tracked items per sentence type.

*Eye movements between animations*

Eye movements during the baseline were first inspected. Table 6.12 shows the mean proportion of looks to target (versus distractor) during the 2500 ms baseline for each age group and sentence type based on participant means. Preference for the target during the baseline ranged between and 46% and 58%.

**Table 6.12 Mean proportion target looks during baseline: Experiment 6a**

| Sentence type | 2½yrs | *sd* | 3½yrs | *sd* |
|---|---|---|---|---|
| + animate + animate | .46 | *.14* | .58 | *.19* |
| + animate - animate | .51 | *.19* | .52 | *.10* |
| - animate + animate | .53 | *.20* | .56 | *.19* |
| - animate - animate | .47 | *.21* | .52 | *.27* |
| Total | .49 | *.10* | .54 | *.13* |

To determine whether S-O animacy depicted in the target animation affected whether it was fixated during the baseline for each age group, the empirical logit transformed mean looks to target were fit to a model with subject animacy, object animacy, and age group as fixed factors, and with participant and item as random factors. There was no significant three-way interaction ($\chi^2(1) = 1.11$, $p > .1$) or two-way interaction ($\chi^2(3) = 1.84$, $p > .1$) between the fixed predictors. Furthermore, there was no main effect of subject animacy ($\chi^2(1) = .15$, $p > .1$), object animacy ($\chi^2(1) = .79$, $p > .1$), or age group ($\chi^2(1) = 1.23$, $p > .1$). Thus, neither the 2½-year-olds nor the 3½-year-olds had a pre-existing preference for either target or distractor in any of the four sentence types. The inclusion of control factors such as gender, test verb, target side, and list showed that target side significantly explained more variance in the data ($\chi^2(1) = 9.06$, $p = .003$), with children more likely to fixate on the target if it was on the left ($\beta = .37$; $t = 2.95$; $p = .003$).

The general pattern of looks to target and distractor over the course of a trial was similar to that of Dutch children, with overall proportions of looks to the target not reaching above .60. Four windows of time were defined for analysis: Time window 1 contains gaze data from the start of the trial to the offset of the sentence subject, with the subject duration about 500 ms. Time windows 2 – 4 are subsequent regions of 1000 ms following the offset of the sentence subject. For each participant and item, the proportion of target looks (versus distractor looks) during each of the four windows was calculated. Figures 6.19 and Figure 6.20 plot the mean proportions of looks to target versus distractor in each of the time windows for each of the age groups.

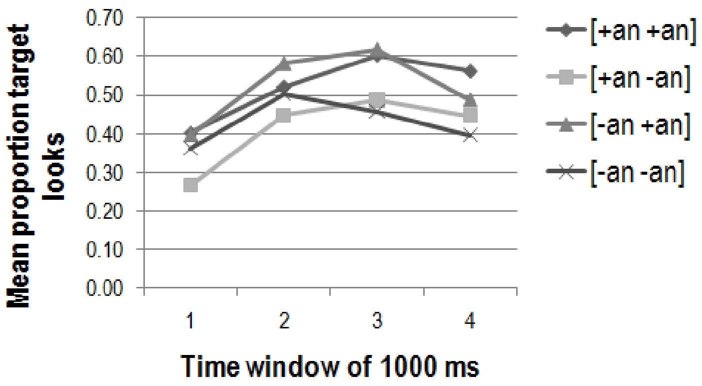


**Figure 6.19 Proportion of looks to target over four time windows: Experiment 6a (English-speaking 2½ yrs)**

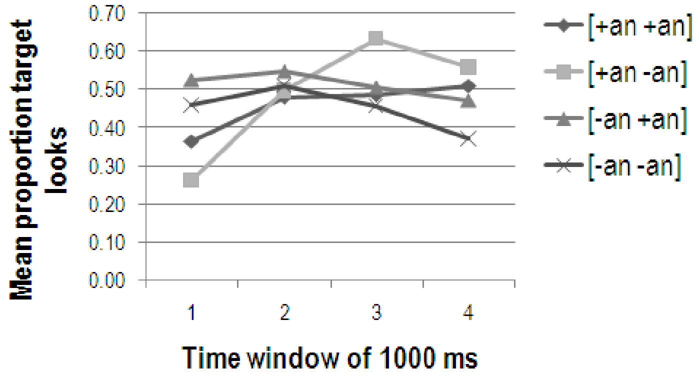


**Figure 6.20 Proportion of looks to target over four time windows: Experiment 6a (English-speaking 3½ yrs)**

To determine whether S-O animacy affected which AOI was fixated during preferential looking of the 2½-year-olds, the empirical logit transformed mean looks to target were fit to a model with subject animacy, object animacy, and time window as fixed factors, and with participant and item as random factors. There was no three-way ($\chi^2(1) = .15$, $p > .1$) interaction between the fixed predictors. There was a two-way interaction between time window and object animacy ($\chi^2(1) = 7.75$, $p = .01$). Together with the main effect of time window ($\chi^2(1) = 31.62$, $p < .001$), this indicates that the younger children looked increasingly towards the target animation as each time window progressed in general ($\beta = .38$; $t = 5.66$; $p < .001$), an effect that was intensified when the object was animate ($\beta = .19$; $t = 2.79$; $p = .01$). The inclusion of control factors in the model of the older children, such as gender, test verb, target side, and list, did not significantly explain more variance in the data.

The same analysis was carried out for the 3½-year-olds. There was no three-way ($\chi^2(1) = 1.34$, $p > .1$) or two-way ($\chi^2(3) = 2.03$, $p > .1$) interactions between the fixed predictors. In the baseline model there was a significant effect of time window ($\chi^2(1) = 23.23$, $p < .001$) with the older children looking increasingly towards the target animation as each time window progressed in general ($\beta = .41$; $t = 4.86$; $p < .001$). Thus, the English-speaking 2½-year-olds' looking reflected a preference for animate objects, whereas the 3½-year-olds showed no such preference. The inclusion of control factors in the model of the older children such as gender, test verb, target side, and list showed that target side significantly explains more variance in the data ($\chi^2(1) = 4.97$, $p < .001$), with participants more likely to fixate on the target if it was on the left ($\beta = .28$; $t = 2.06$; $p = .04$). Target side appears not to interact with time window ($\chi^2(1) = .03$; $p > .10$) as it had for English-speaking adults.

*Eye movements within animations*

AOIs were further defined within target and distractor animations as either the agent of the action or the patient of the action. The agent advantage score within either the target or distractor animation, for each sentence type, based on participant means are listed for each age group in Table 6.13 and Table 6.14. The mean agent advantage score for sentence types with no S-O contrast (S=O) are shown in Figure 6.21 and for sentence types with an S-O contrast (S≠O) in Figure 6.22.

**Table 6.13 Mean agent advantage score for target and distractor over trial: Experiment 6a (English-speaking 2½ yrs)**

| Sentence type | Agent advantage score **within target** | *sd* | Agent advantage score **within distractor** | *sd* |
|---|---|---|---|---|
| + animate + animate | .23 | *.30* | .01 | *.44* |
| + animate - animate | .33 | *.26* | -.26 | *.26* |
| - animate + animate | -.17 | *.35* | .29 | *.36* |
| - animate - animate | .12 | *.33* | .20 | *.38* |
| Total | .12 | *.14* | .03 | *.17* |

**Table 6.14 Mean agent advantage score for target and distractor over trial: Experiment 6a (English-speaking 3½ yrs)**

| Sentence type | Agent advantage score **within target** | *sd* | Agent advantage score **within distractor** | *sd* |
|---|---|---|---|---|
| + animate + animate | .13 | *.37* | .18 | *.44* |
| + animate - animate | .26 | *.43* | -.31 | *.35* |
| - animate + animate | -.21 | *.37* | .23 | *.40* |
| - animate - animate | .01 | *.49* | .05 | *.43* |
| Total | .02 | *.13* | .06 | *.20* |

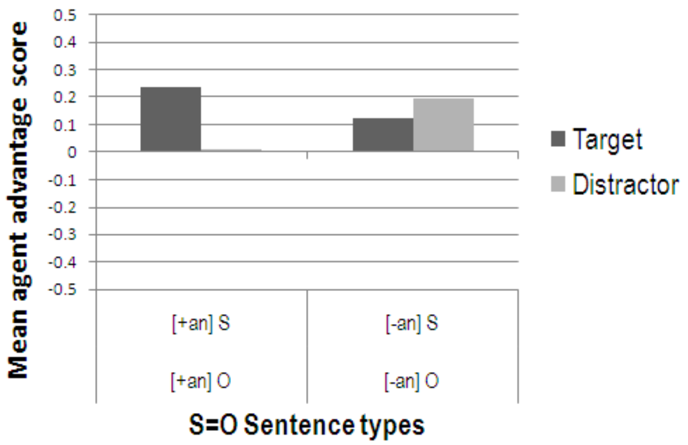


**Figure 6.21 Mean agent advantage score for S=O sentences: Experiment 6a (English-speaking preschoolers)**

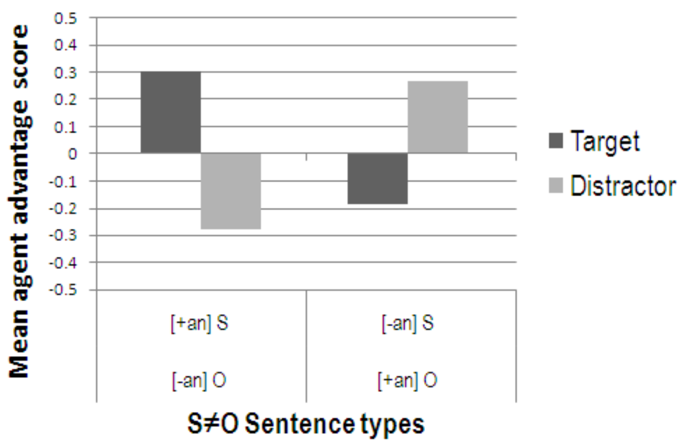


**Figure 6.22 Mean agent advantage score for S≠O sentences: Experiment 6a (English-speaking preschoolers)**

To determine whether agent animacy affected the degree to which the agent was fixated in either target or distractor animation during the trials with S=O sentences by either age group, the empirical logit transformed agent advantage scores for each participant and item were fit to a model with agent animacy, animation (target vs. distractor), and age group as fixed factors, and with participant and item as random factors. There was no three-way ($\chi^2(1) = .25$, $p > .1$) or two-way ($\chi^2(1) = 1.67$, $p > .1$) interaction between the fixed predictors, so main effects were checked for in the baseline model. There was no effect of animation ($\chi^2(1) = .11$; $p > .1$), agent animacy ($\chi^2(1) = .01$, $p > .1$), or age group ($\chi^2(1) = .48$, $p > .1$). Thus, the children showed positive mean agent advantage scores on average across the animation types and sentence types, but there was no discernable effect of any of the factors.

The same analysis was carried out for trials with S≠O sentences. There was no three-way interaction between the fixed predictors ($\chi^2(1) = .4$, $p > .1$), but there was a significant two-way interaction between agent animacy and animation ($\chi^2(1) = 69.51$; $p < .001$). Thus, the English-speaking children were more likely to fixate the agent that was within the target animation only in the [+an -an] sentences, in which the agent was animate ($\beta = 2.15$; $t = 8.72$; $p < .001$).

### 6.5.3 Discussion

Results of the picture selection task in Experiment 6a show that English-speaking preschoolers' interpretations did not vary as a function of the manipulation of either subject or object animacy. There was an effect of age, with the younger children correctly interpreting word order 60% of the time on average and the older children 80%. The children also chose the target more often when it was on the right, which may reflect a right hand bias when pointing. These findings are not in line with previous studies that showed English-speaking preschoolers at age 2;6 exhibit variable performance on word order in the face of a conflict with animacy information. Consequently, these results do not meet the predictions of the OT model based on the variable performance of English-speaking preschoolers found in previous studies.

Results of the preferential looking task in Experiment 6a show that the English-speaking 3½-year-olds' gaze did not vary as a function of the manipulation of animacy, which is in line with the predictions. These older children were expected to rely solely on word order when interpreting the sentences. It was the 2½-year-olds for which it was expected that S-O animacy would play a role. Indeed, there was an effect of object animacy, but it was in the opposite direction than expected. These younger children showed a greater increase in looks to the target in the sentences with objects that were animals than in the sentences with objects that were vehicles. It is not clear why the two [+an] object sentence types would earn greater target looks than the [-an] object sentence types since it is proposed that inanimate entities are easier to interpret as objects than animate entities. We can assume that this unexpected effect was due to the linguistic input and not pre-existing preferences for pictures in which animals are patients (not that this would be expected) since the 2.5 second baseline revealed equal looking to target and distractor across sentence types. As in Experiment 4a with Dutch preschoolers, the baseline looking behavior was likely a scanning from left to right, evidenced by the significant effect of target side.

The analysis of within-animation looks by English-speaking children and adults showed a pattern that paralleled that of the Dutch adults and children. The English-speaking adults showed the same preference as Dutch adults for the agent in the target animations when there was no contrast in animacy between the entities in the animations. The English-speaking preschoolers showed no significant preference for any entity within S=O target or distractor animations, but did have overall positive mean agent advantage scores, indicating that looks were generally towards agents. Regarding the animations in which there was a contrast in animacy between the entities in the animations, both English-speaking adults and preschoolers showed a preference for the agent in the target picture only in the [+an -an] sentences. This shows that the preference to look at the agent of the action except for in S<O actions during this type of comprehension task is one that holds across Dutch and English.

## 6.6 General Discussion

The questions that Experiments 3a – 6a aim to answer is (i) whether Dutch- and English-speaking preschoolers exhibit variable comprehension of word order due to S-O animacy that adults do not, and (ii) whether Dutch- and English-speaking adults show processing effects due to S-O animacy. The results of Experiments 3a – 6a, which tested sentences with an animal-vehicle distinction, are summed up as follows:

- Experiment 3a tested the comprehension of 41 Dutch adults and Experiment 5a tested the comprehension of 31 English-speaking adults using a picture selection task with animations.
  - Adults of both languages were more likely to choose the animation representing an SO interpretation when the subject was animate than when the subject was inanimate.
  - Adults of both languages were about 200 ms – 300 ms faster to choose animations representing SO interpretations when the subject was animate than when the subject was inanimate.
  - Adults of both languages looked increasingly towards the target animation before making their choice. The Dutch adults showed a greater increase of looking to the target animation when the subject was animate than when the subject was inanimate, an effect that was not found in English-speaking adults.

- Experiment 4a tested the comprehension of 32 Dutch preschoolers using a picture selection task and preferential looking task with animations.
  - Both 2½-year-olds and 3½-year-old Dutch preschoolers interpreted word order correctly when S>O (72% – 79%), and they interpreted word order with difficulty when S<O (48% – 54%).
  - In the preferential looking task, children of both age groups looked increasingly towards the target animation as time progressed. The Dutch 3½-year-olds showed a greater increase of looking to the target animation when the subject was animate than when the subject was inanimate, an effect that was not found in the Dutch 2½-year-olds.

- Experiment 6a tested the comprehension of 31 English-speaking preschoolers using a picture selection task and preferential looking task with animations.
  - In the picture selection task, English-speaking preschoolers exhibited no effect of S-O animacy in either age group. Across sentence types, the younger children relied on word order about 60% of the time and the older children about 80%.
  - In the preferential looking task, children of both age groups looked increasingly towards the target animation as time progressed. The English-speaking 2½-year-olds showed a greater increase of looking to the target animation when the object was animate than when the object was inanimate.

In the following discussion, these results are interpreted in light of the questions asked. First, offline responses are considered, and then processing effects are discussed. Finally, the significance of the within-animation gaze patterns is addressed.

*Effect of animacy on interpretation*

Do Dutch- and English-speaking preschoolers exhibit variable comprehension of word order due to S-O animacy that adults do not? To answer this first question, we can use predictions for offline behavior from Hendriks et al.'s (2005) model of S-O word order interpretation for preschoolers. Their model predicts that a child's grammar with a too highly ranked animacy constraint does not find an SO interpretation optimal for S<O sentences, whereas an adult's grammar with the correct ranking of word order over animacy will find SO interpretations optimal for all sentence types. These predictions hold for both Dutch and English. Variable comprehensions stemming from mis-ranked constraints is expected in children of both languages at age 2;6, but only in Dutch children still at the age of 3;6.

For Dutch preschoolers, based on the results of the picture selection task, both the 2½-year-olds and the 3½-year-olds were influenced by S-O animacy when interpreting sentences. Worst performance for both groups was on S<O sentences, which is precisely in line with predictions. The results of the preferential looking task show that both groups preferred to look at the target over the distractor as the trial progressed, indicating a general understanding of word order by Dutch preschoolers. It was only the older group of Dutch children who showed an increased preference for the target when the subject was animate versus when the subject was inanimate. These preferential looking results are similar to those found in Experiment 2a, in which only the older group of Dutch children revealed an effect of animacy.

For English-speaking preschoolers, based on the results of the picture selection task, neither the 2½-year-olds nor the 3½-year-olds were influenced by S-O animacy when interpreting sentences. The results of the preferential looking task show that both groups preferred to look at the target over the distractor as the trial progressed, indicating a general understanding of word order by English-speaking preschoolers. It was only the younger group of English-speaking children who showed an increased preference for the target when the object was animate versus when the object was inanimate. The only aspect of these results that meet the predictions made is that the 3½-year-olds showed no effect of animacy. The fact that the 2½-year-olds did not show an effect of animacy in the picture selection task and that they showed an effect of animacy in the preferential looking task that went in the opposite direction than expected goes against the predictions made.

In Experiments 4a and 6a with preschoolers of Dutch and English, there was no significant effect of gender in the responses of the children in either the picture selection task or the preferential looking task. It was suggested in Chapter 4 that boys may find toy vehicles more salient than girls, a confounding factor that may have conflicted with any effect of the animacy manipulation. The fact that there was no effect of gender in this study with only animated depictions of vehicles can be seen as evidence that the gender effect is associated with the task with tangible toy vehicles.

For adult speakers of Dutch and English, there were animacy effects found in the offline measure. Although they were not predicted, these effects are compatible with the OT model. There was an effect of subject animacy that was significant in an overall set of responses that, on average, reflected an SO interpretation 97% of the time. This can be seen as strong evidence for an animacy constraint that is present in the grammar of Dutch- and English-speaking adults. Participants behaved as if they needed to answer as soon as possible on the simple task, often giving answers before the end of the sentence, even though they were instructed to answer as quickly and accurately as possible. This might explain why the underlying presence of an animacy constraint was evident even in the offline answers the adults gave.

*Effect of animacy on processing*

Do Dutch- and English-speaking adults show processing effects due to S-O animacy? To answer this second question, we can use predictions of my incremental model that follows Lamers and de Hoop's

(2004) OT model of S-O word order interpretation for adults; the model predicts that sentences with inanimate subjects, especially S<O sentences, are at a disadvantage during processing. Such effects were found in the reaction time measures of the adults of both languages, who were about 200 ms – 300 ms faster to choose animations reflecting SO interpretations when the subject was animate. The Dutch adults also showed a greater increase of looking to the target when the subject was animate, an effect that was not found in English-speaking adults. Since Dutch allows for more flexible word order, one might expect greater effects of animacy in Dutch. That is, since Dutch listeners expect that information besides word order may be used (agreement, case, verb selection properties), they might be more influenced by animacy information in the visual scene than English-speaking adults, even if the animacy constraint is proposed to be lower ranked in their grammar than the word order constraint.

*Within-animation gaze*

So far, gaze results have been discussed in terms of a competition between target and distractor. How was animacy information in the visual scene used in terms of within-animation gaze patterns? There was a pattern of within-animation gaze that held more or less (barring some slight variation in the preschoolers' gaze for no-contrast animation pairs) in which participants looked at the agent in animations when there were all animals or all vehicles in the animations. The preference to find the agent was disrupted when there was an animacy contrast in agents and patients in the target and distractor pictures. In this case, the participants were likely to fixate on animate entities, regardless of whether the animate entity was agent or patient.

When interpreting the within-animation gaze behavior, we can make use of the assumption, introduced in Section 3.1.2, that eye movements during comprehension are mediated by the linguistic task (Eberhard et al., 1995). That is, people look where they need to to carry out the task. Furthermore, where people look when presented with different types of visual scenes may differ when processing the same type of syntactic structure (Spivey et al., 2002). In this study, participants looked longest to the agents within each animation when all entities were equal in animacy, which can be interpreted as a linguistically mediated search for the agent. When the entities in each animation were not equal in animacy, participants looked longest to the animate entities within each animation. If this preference to look at animals is interpreted as an extension of the search for agent, then the behavior seems to be evidence for an animacy constraint that favors animate agents and inanimate patients. However, it could be argued that people simply prefer to look at things that are alive, regardless of their role in an action. The gaze patterns during sentence production presented in Chapter 7 may shed light on which of these is the case since gaze during production is also mediated by the linguistic task (Griffin & Bock, 2000; Papafragou et al., 2008). If participants do not prefer to look at the alive entity when viewing a single animation during a production task, this could be seen as support for a preference for animate agents and inanimate patients that occurs in comprehension only.

In sum, both adults and children, save for the English-speaking preschoolers, showed effects of S-O animacy that are in line with the existence of an animacy constraint in their grammars: a preference for animate subjects as well as preference for inanimate objects This is based on both the interpretations they had for the sentences, as well as how they processed the sentences—either in their speed of response or the way they visually inspected the two choices they were presented with. In order to see whether S-O animacy plays a role in comprehension, but not production, the same populations were tested in a production task, presented in the next chapter.

# 7 Production: Elicitation with animations

Experiments 3b, 4b, 5b & 6b

This chapter presents production experiments in which Dutch- and English-speaking preschoolers are tested with a sentence elicitation task with cartoon animations as stimuli. Dutch- and English-speaking adults are additionally tested as controls. Four production experiments are reported in this chapter, one for each of the four populations tested. The accompanying comprehension experiments carried out with the same populations and materials were presented in the previous chapter. (An overview of all experiments is provided in Table 1.1 at the end of Chapter 1.)

## 7.1 Introduction

The few studies that have tested preschoolers on sentences in which S-O animacy and word order are pitted against each other found that the utterances elicited by English-speaking children adhere to the word order rules of their language (Angiolillo & Goldin-Meadow, 1982; Chapman & Miller, 1975; McClellan et al., 1986). The same was found for the Dutch preschoolers (above 80% SO order) tested in the act-out tasks in Experiments 1b and 2b. The experiments presented in this chapter test word order production using the same sentences that were tested in Experiment 1, which used an animal-vehicle animacy distinction. Rather than describing actions they see carried out with toys, participants describe actions they see depicted in cartoon animations.

The first question that Experiments 3b – 6b aim to answer is whether adult-like production of S-O word order is found in a well-controlled sentence elicitation task with animations carried out with Dutch- and English-speaking preschoolers. Hendriks, de Hoop, and Lamers' (2005) model of S-O word order in preschoolers predicts adult-like production of word order in both Dutch- and English-speaking preschoolers. If children's production of word order is better than their comprehension of word order in more than one type of comprehension-production task pair, this would be seen as strong evidence that there is a developmental asymmetry.

The second question these experiments aim to answer is whether animacy affects production. Hendriks et al.'s account proposes that the early grammar initially gives too much priority to an animacy constraint, but the animacy constraint is one that affects meaning only. Therefore, the mis-ranking does not affect word order production. By testing adult control groups, Experiments 3b and 5b test the predictions of the model of adult grammar proposed by Hendriks et al. (2005), namely that adults of English and Dutch will rely on word order over animacy in sentence production.

Regarding the pattern of eye movements during elicitation, there are expectations that follow from the results of Griffin and Bock (2000) with English-speaking adults. It is expected that both the Dutch- and English-speaking adult and child populations in the present studies will exhibit an agent-then-patient gaze pattern during the production of simple, transitive, active sentences. That is, the speaker first looks to the agent of the action in the scene (which will be the subject of their active sentence), and they begin to speak as they look to the patient of the action in the scene (which will be the object of their active sentence).

The third question is whether animacy will affect the gaze behavior of the speakers in the experiments. It was discussed in Section 2.3.1.2 that there is a cross-linguistic preference for speakers to place animate entities, assumed to be highly accessible, early in their sentences (Branigan et al., 2008; van Bergen, 2011). The distinction was emphasized between contextual accessibility of animate entities, brought on by their tendency to be prominent in the discourse, and conceptual accessibility, resulting from the inherent availability of animate entities at the conceptual level. Since the sentences are spoken out of context, only the inherent accessibility intrinsic to being either an animal (versus a lesser accessible vehicle) remains and is not expected to trigger word order alternations. However, animations with animate agents may experience some facilitation of the agent-then-patient pattern during production, due to the inherent accessibility of animate entities.

## 7.2 Experiment 3b

### 7.2.1 Method

*Participants*

The same Dutch adults ($n$ = 41) who participated in the picture selection task in Experiment 3a also participated in an elicitation task.

*Design and materials*

In the production task, sentences were elicited from participants using target animations from the comprehension task in 3a, so that the exact same sentences were tested in both tasks.

*Apparatus*

The computer set-up was the same as in Experiment 3a except that participants heard no audio, but rather produced sentences that were recorded on a microphone via E-Prime.

*Procedure*

Each participants' eyes were first calibrated at five points on the screen. Then the elicitation task was administered either before or after the picture selection task described in Experiment 3a. Participants were instructed to describe the animation that appeared on the screen in a short sentence, with three chances to practice and receive feedback if necessary. They were given the example that if they saw a picture of a panda kissing a lion, they should say "The panda is kissing the lion." They were also instructed not to say the Dutch equivalent of the construction "There is/are" since usually there was a clear action occurring in the pictures. After the participant had produced a sentence, the experimenter pressed a button that advanced them to the next item. In order to give the participants time between the sentence productions, a fixation point with duration of two seconds was programmed to precede each animation.

*Scoring*

As in Experiments 1b and 2b, utterances were categorized as scorable (*SO order* or *OS order*), or unscorable (*Nontarget, Insufficient*, or *Unclear*). An additional category of *Passive* was included for sentences in the passive form. If a revision was made, the final utterance was used for scoring. An independent transcription of 10% of the participants ($n$ = 4) showed 91% agreement with the main transcription. Inter-scorer agreement was also high (Cohen's κ = .94). Voice onset latency (VOL), or how long the participant viewed the animation before beginning to describe it, was measured for each item.

### 7.2.2 Results

*Criteria for inclusion*

There were 3 participants excluded because they did not contribute at least two scorable responses per sentence type, attributed to too many *Nontarget* responses. The removal of these participants decreased the number of participants in the analysis to 38.

*Accuracy*

Of the 608 responses given by Dutch adults, 577 were scorable and 29 were unscorable (all *Nontarget*), and 2 were *Passive*. Of the scorable responses, adults used *SO order* 100% of the time.

*Voice onset latency*

Items with extreme VOLs ($n$ = 5) were removed from the VOL analysis. Extreme VOLs were defined as those outside 3 standard deviations of the participant's personal mean. Mean VOLs on the four sentence types based on participant means are shown in Table 7.1 and Figure 7.1. It took the adults 1402 ms on average to begin a sentence.

**Table 7.1 Mean VOL for producing SO order: Experiment 3b (Dutch adults)**

| Sentence type | Mean | *sd* |
|---|---|---|
| + animate + animate | 1422 | *341* |
| + animate - animate | 1267 | *278* |
| - animate + animate | 1518 | *437* |
| - animate - animate | 1418 | *355* |
| Total | 1402 | *313* |

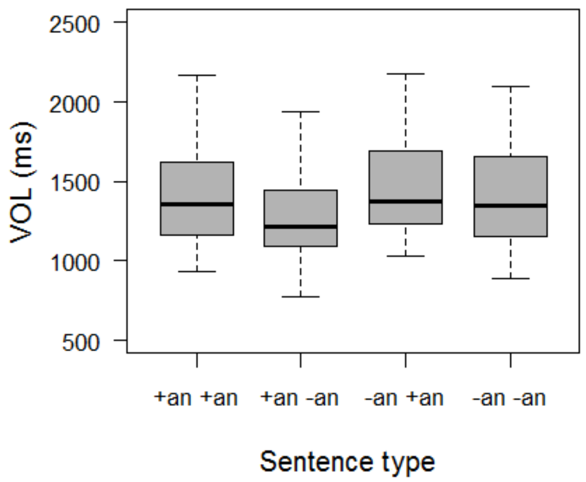


**Figure 7.1 Mean VOL for producing SO order: Experiment 3b (Dutch adults)**

To determine whether S-O animacy affected speed of response in the elicited production task, the log transformed VOLs were fit to a model with subject and object animacy as fixed factors and participants and items as random factors. Since including an interaction was not justified ($\chi^2(1) = .77$, $p > .1$), the baseline model was checked for main effects. There was a significant main effect of both subject animacy ($\chi^2(1) = 6.03$, $p = .01$) and object animacy ($\chi^2(1) = 4.80$, $p = .03$). The adults had longer VOLs when the subject was inanimate ($\beta = .04$; $t = 2.49$; $p = .01$) as well as when the object was animate ($\beta = .04$; $t = 2.22$; $p = .03$). The inclusion of control factors such as test verb, first task, direction of action, and list showed that first task significantly explained more variance in the data ($\chi^2(1) = 9.57$, $p = .002$). Dutch adults had longer VOLs when they performed the elicitation task before the picture selection task ($\beta = .10$; $t = 3.20$; $p = .001$). Thus, the Dutch adults were slower to begin their sentences when the subject was inanimate as well as when the object was animate. In addition, they were able to begin their sentences faster in general when they were more familiar with the animations and sentences (i.e. had encountered them previously in the comprehension task).

*Eye movements*

Items that remained in the VOL analysis were used in the analysis of eye movements. There was 1 adult excluded because most of her trials had extreme track loss. An item had extreme track loss if there was track loss of both eyes for more than one-third of the four-second region of interest. Remaining items with extreme track loss ($n = 8$) were also removed.

Areas of Interest (AOIs) in the visual stimuli were defined over *Agent*, *Patient*, and *Not on AOI*. Figure 7.2 shows the general pattern of looks to agent and patient over the course of a trial, synchronized to the onset of each participant's sentence. The gaze plot shows that Dutch adults looked first to the agent (subject) prior to starting a sentence, and then to the patient (object).

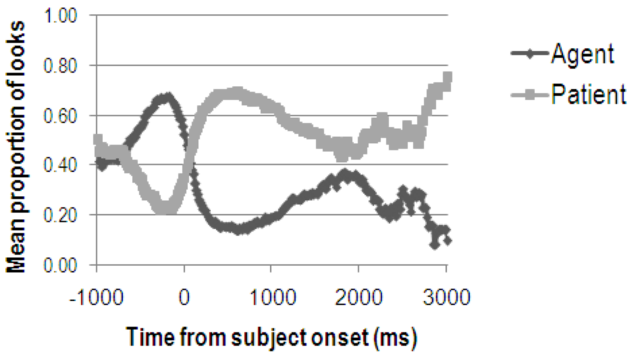


**Figure 7.2 Proportion of looks to agent and patient: Experiment 3b (Dutch adults)**

Two windows of time were defined for analysis: Time window 1 (pre-subject) contains gaze data during 1000 ms prior to the onset of the sentence subject, and Time window 2 (post-subject) contains gaze data during 1000 ms after the onset of the sentence subject. For each participant and item, the agent advantage score was calculated—that is, the difference between the proportion of looks to the agent minus the proportion of looks to the patient during each of the two windows. Figure 7.3 plots the mean agent advantage score in each time window. Negative scores occur in the second time window, indicating an advantage of the patient after the participants have begun their sentences. Note that the plots of agent advantage score in the previous chapter differ from these plots in that those in the previous chapter plotted a global score over the trial (represented by a bar plot), while these plot a score that changes over time (represented by a line graph).

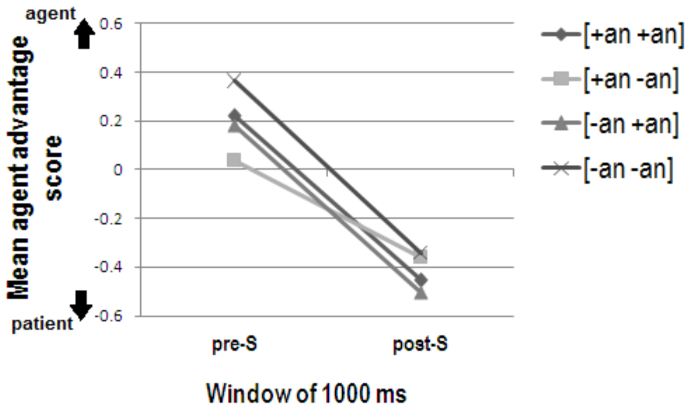


**Figure 7.3 Mean agent advantage score over two time windows: Experiment 3b (Dutch adults)**

To determine whether S-O animacy affected which AOI was fixated during sentence planning and production, the empirical logit transformed mean agent advantage scores from each time window were fit to a model with subject animacy, object animacy, and time window as fixed factors, and with participant and item as random factors. There was a significant three-way interaction between the fixed predictors ($\chi^2(1) = 5.89$, $p = .02$), which could be interpreted in light of a significant effect of time window ($\chi^2(1) = 260.69$, $p < .001$). In all sentence types, there was a decrease in the preference for agent over patient from the first to the second time window ($\beta = -4.16$.; $t = 17.18$; $p < .001$), but to a significantly lesser degree in the [+an -an] sentences ($\beta = -.58$.; $t = -2.43$; $p < .02$). Thus, Dutch adults increased their looks to patient as their sentence unfolded, but to a lesser degree for [+an -an] sentences. The inclusion of control factors such as test verb, first task, direction, and list showed that the inclusion of verb x time window significantly explained more variance in the data ($\chi^2(1) = 94.44$, $p < .001$), with greater looks to the agent over patient in the first time window when the verb was *pull* ($\beta = 4.26$.; $t = 9.01$; $p < .001$).

### 7.2.3 Discussion

Results of Experiment 3b show that Dutch adults produced SO word order 100% of the time when describing the animations. Their VOLs, on the other hand, varied as a function of the manipulation of subject animacy as well as of object animacy. Dutch adult started their sentences more quickly when the subject was animate and when the object was inanimate. They were also faster overall to produce sentences when they received the production task after the comprehension task described in Experiment 3a. This makes sense, as the participants who received production second were more familiar with the animations and sentences than those who received production first.

The gaze pattern of the Dutch adults as they produced the simple, transitive sentences reflects the pattern found by Griffin and Bock (2000) for English-speaking adults. For all sentence types, the Dutch adults looked first to the agent (the subject of the active sentence they produced) and began their sentence as they switched their gaze to the patient (the object of the active sentence they produced). The S>O sentences saw the smallest decrease in looks to the agent in favor of the patient during one second after the sentence was started. If taken together with the fact that the combination of subject animacy and object inanimacy in these sentences gave them the greatest advantage regarding speed of response (VOL), it could be that only brief looks to the agent and patient were necessary in order to create the utterance. These results can be interpreted as a facilitation of word order production due to inherent accessibility of animate entities.

In sum, the Dutch adults met the predictions that they would use SO order across the sentence types. They also met the prediction that they would exhibit an agent-then-patient gaze pattern during sentence production. Furthermore, the prediction was met that sentences with animate subject would be facilitated during production (found in the speed of VOL and eye movements), in this case specifically for S>O sentence types. Thus, the inherent accessibility of animate entities resulted in quicker responses, but not word order alternations. We now turn to Dutch preschoolers to see how they fare in the same production task.

## 7.3 Experiment 4b

### 7.3.1 Method

*Participants*

The same Dutch 2½-year-olds ($n$ = 15) and 3½-year-olds ($n$ = 17) who participated in the picture selection and preferential looking tasks in Experiment 4a also participated in an elicitation task.

*Design and materials*

The 16 test items (both animations and sentences) were identical to those in Experiment 4a, Similar practice items were used (listed in Table A.6 in Appendix A.), but no fillers were included. The direction of action of the animations was balanced across the four sentence types. In the lists, no entity appeared in the preceding or following item, nor were adjacent items of the same sentence type.

*Apparatus*

The computer set-up was the same as in Experiment 3b except that productions were recorded via a separate voice recorder instead of via E-Prime.

*Procedure*

The production procedure immediately followed the comprehension procedure described in Experiment 4a, which had been in turn preceded by a toy naming and verb demonstration pre-test. The elicited production task always immediately followed the preferential looking task and tested the same verb. The task, described below, was preceded by three practice items.

To begin, a hand puppet was introduced: Joris the sheep. The assistant explained to the child that Joris would close his eyes so that he could not see what was happening. It was the child's task to tell Joris what was happening in the animation. The phrase *What's happening?* was used instead of *What do you see?* in order to encourage an answer that described the entire action. Joris would then open his eyes and check to see if the child was right. Joris always gave positive praise except in the case of no answer, in which case he would encourage the child to try again on the next item. The same small break and toy reward described in Experiment 4a was shared by this elicitation experiment.

*Scoring*

The same scoring procedure was employed as in all previous "b" Experiments. An independent transcription of 10% of the participants ($n$ = 3) showed 90% agreement with the main transcription. Inter-scorer agreement was also high (Cohen's $\kappa$ = .90).

For each child's scorable utterance, the moment at which they began their sentence was measured. Because the children generally took much longer than the 1 – 2 seconds the adults took to begin a sentence and sometimes had to be reminded to describe the picture during this time by the assistant, this measure was not considered to be a VOL. Instead, it was a measure used for the synchronization of the gaze data.

### 7.3.2 Results

*Criteria for inclusion*

There were 11 children who were excluded because they did not contribute at least two scorable responses per sentence type (2 of whom had also been excluded from the pointing accuracy analysis in 4a for the same reason). The removal of these participants decreased the number of 2½-year-olds by 10 (*n* = 5, *mean* = 2;11, *sd* = 3.5 mo.s) and 3½-year-olds by 1 (*n* = 16, *mean* = 3;8, *sd* = 2.9 mo.s).

*Scorability*

The utterances produced by the remaining children were largely scorable. Of the 80 responses given by 2½-year-olds, 70 were scorable and 10 were unscorable*:* 5 *Nontarget*, 3 *Insufficient*, and 2 *Unclear*. Of the 256 responses offered to 3½-year-olds, 232 were scorable and 24 were unscorable (all *Nontarget*).

*Accuracy*

Mean accuracy for each group on the four sentence types based on participant means are shown in Table 7.4 and Figure 7.4. On average, the 2½-year-olds used *SO order* 81% of the time (and 19% for *OS order*), while the 3½-year-olds used *SO order* 92% of the time (and *OS order* 8%). (Individual scores can be found in Table C.13 and Table C.14. in Appendix C.)

**Table 7.4 Mean proportion SO order produced: Experiment 4b (Dutch preschoolers)**

| Sentence type | 2½yrs | *sd* | 3½yrs | *sd* |
|---|---|---|---|---|
| + animate + animate | 1.00 | - | .93 | *.18* |
| + animate - animate | .90 | *.22* | .96 | *.17* |
| - animate + animate | .60 | *.25* | .92 | *.18* |
| - animate - animate | .73 | *.28* | .88 | *.21* |
| Total | .81 | *.13* | .92 | *.17* |

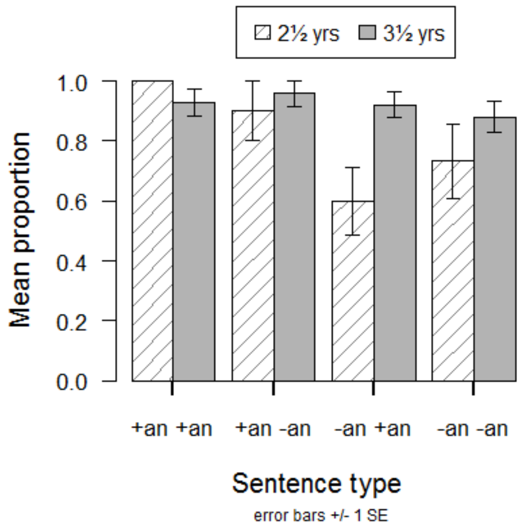


**Figure 7.4 Mean proportion SO order produced: Experiment 4b (Dutch preschoolers)**

To determine whether S-O animacy affected accuracy in production in either age group, the binomial data were fit to a model with subject animacy, object animacy, and age group as fixed factors and participants and items as random factors. The fully complex model was not fit due to complete collinearity between the three-way interaction of the fixed predictors and the two-way interaction age group and subject animacy. The three-way interaction was therefore not included in the model as a strategy to reduce this severe collinearity (Baayen, 2008: 183). There were no significant two-way interactions ($\chi^2(3) = 1.93$, $p > .1$). In the baseline model, there was no main effect of object animacy ($\chi^2(1) = 1.94$, $p > .1$). Subject animacy was a significant predictor ($\chi^2(1) = 8.87$, $p = .003$), with the children more likely to use SO order when the subject was animate ($\beta = .81$; $z = 2.97$; $p = .003$). An effect of age group that approached significance ($\chi^2(1) = 3.47$, $p = .06$) indicated that the older children tended to use more SO order than the younger children ($\beta = 1.20$; $z = 1.86$; $p = .06$). The inclusion of control factors such as gender, test verb, direction of action, list and vocabulary score showed that both verb ($\chi^2(1) = 12.20$, $p < .001$) and direction ($\chi^2(1) = 4.63$, $p = .03$) significantly explain more variance in the data, with the children more likely to use SO order when the verb was *push* ($\beta = .90$; $z = 3.18$; $p = .001$) as well as when the direction was to the left ($\beta = .54$; $z = 2.05$; $p = .04$).

*Eye movements*

All of the children included in the accuracy analysis were included in the production gaze analysis except for one older child who had too many trials across conditions with extreme track loss. An item had extreme track loss if there was track loss of both eyes for more than one-thirds of the three-second window of interest. Test items with extreme track loss ($n = 56$) as well as all OS responses were removed from the analysis of eye movements. Each child contributed at least two validly tracked items per sentence type, except for 6 children who contributed at least two valid items three sentence types and 1 child who contributed at least two valid items on two sentence types.

AOIs in the visual stimuli were defined over *Agent*, *Patient*, and *Not on AOI*. Figure 7.5 and Figure 7.6 show the general pattern of looks to agent and patient over the course of a trial for each age group, synchronized to the onset of each child's sentence. The gaze plot shows that Dutch preschoolers looked first to the agent (subject) prior to starting a sentence, and then to the patient (object). The crossover from agent to patient was protracted from subject onset by about 250 ms by the older group and about 750 ms by the younger group, of which there were only 5 children.

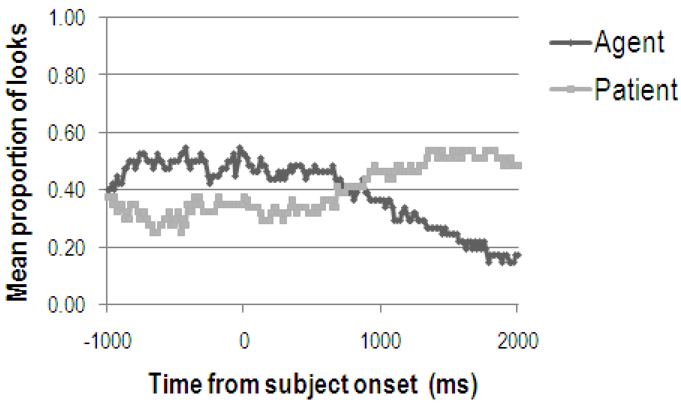


**Figure 7.5 Proportion of looks to agent and patient: Experiment 4b (Dutch 2½ yrs)**

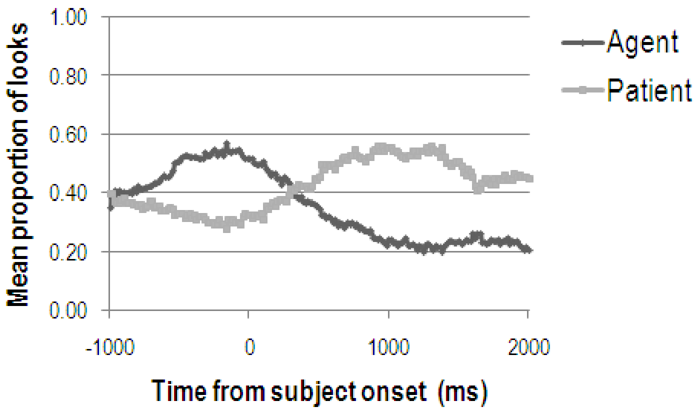


**Figure 7.6 Proportion of looks to agent and patient: Experiment 4b (Dutch 3½ years)**

Two windows of time were defined for analysis: Time window 1 contains gaze data during 1000 ms prior to the onset of the sentence subject, and Time window 2 contains gaze data during 1000 ms after the onset of the sentence subject. For each participant and item, the agent advantage score was calculated. Figure 7.7 and 7.8 plots the mean agent advantage score in each time window for each age group. Negative scores indicate an advantage of the patient.

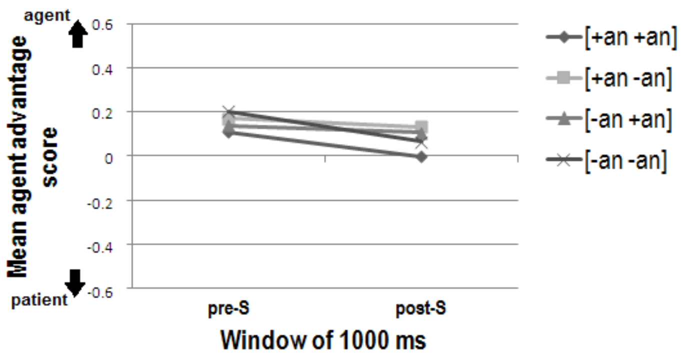


**Figure 7.7 Mean agent advantage score over two time windows: Experiment 4b (Dutch 2½ yrs)**

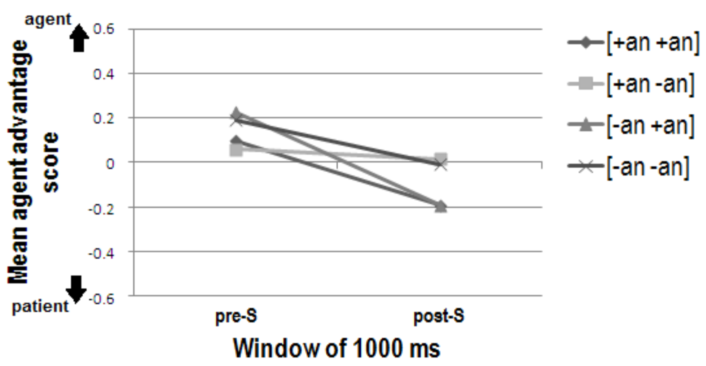


**Figure 7.8 Mean agent advantage score over two time windows: Experiment 4b (Dutch 3½ yrs)**

To determine whether S-O animacy affected which AOI was fixated during sentence planning and production for each age group, the empirical logit transformed mean agent advantage scores from each time window were fit to a model with subject animacy, object animacy, time window, and age group as fixed factors, and with participant and item as random factors. There were no significant four-way ($\chi^2(1)$ = .07, $p$ > .1), three-way ($\chi^2(4)$ = 1.30, $p$ > .1), or two-way ($\chi^2(6)$ = 5.57, $p$ > .1) interactions between the fixed predictors. In the baseline model there was no effect of subject animacy ($\chi^2(1)$ = 1.10; $p$ > .1) or object animacy ($\chi^2(1)$ = .31; $p$ > .1). There was a significant main effect of time window ($\chi^2(1)$ = 16.51; $p$ < .001), with the agent advantage decreasing from the first to the second time window ($\beta$ = -1.68; $t$ = -4.09; $p$ < .001).

To be sure, a second analysis was run with only the older group of children, thereby excluding the five younger children. Again, there were no significant three-way ($\chi^2(1)$ = .03; $p$ > .1), or two-way ($\chi^2(3)$ = 3.93; $p$ > .1) interactions between the fixed predictors. An interaction of time window and object animacy, however, approached significance ($\chi^2(1)$ = 3.16; $p$ = .08). This, in combination with the significant main effect of time window ($\chi^2(1)$ = 503.21; $p$ < .001) means that there was a decrease in agent advantage score from the first to the second time window ($\beta$ = -1.95; $t$ = -4.37; $p$ < .001), a decrease that tended to be even greater for the sentences with an animate object ($\beta$ = -.79; $t$ = -1.77; $p$ = .08).

The inclusion of control factors such as gender, test verb, direction of action, and list showed that the test verb significantly explained more variance in the data. For both the model with the younger children ($\chi^2(1)$ = 12.17; $p$ < .001) and the one without ($\chi^2(1)$ = 12.88; $p$ < .001), there was greater looking to the agent when it was pulling: ($\beta$ = 1.11; $t$ = 3.51; $p$ < .001) and ($\beta$ = 1.05; $t$ = 3.67; $p$ < .001).

### 7.3.3 Discussion

Results of Experiment 4b show that S-O word order used by Dutch preschoolers varied as a function of the manipulation of animacy. The mean SO word order used by the children in both age groups when the agent was inanimate was lower than when the agent was animate. Such an effect of subject animacy was not predicted on the basis of Hendriks et al.'s (2005) OT model. While there was no effect of animacy found in the elicitation tasks in Experiments 1b and 2b with Dutch preschoolers, the overall use of SO in those experiment are comparable to that of Experiment 4b In Experiments 1b and 2b, the 2½-year-olds used SO word order 84% – 91% of the time and the 3½-year-olds 95% – 99% of the time. In Experiment 4b, the 2½-year-olds produced SO word order 81% of the time 3½-year-olds 92% of the time;

The gaze pattern of the Dutch preschoolers was adult-like in that it reflected the pattern found by Griffin and Bock (2000) for simple, transitive sentences. For all sentence types, the children looked first to the agent (the subject of the active sentence they produced) and began their sentence as they switched their gaze to the patient (the object of the active sentence they produced). The older children began a sentence roughly 250 ms after they had viewed the agent, while the younger children began a sentence roughly 750 ms after they had viewed the agent. The younger children were a small group and very protracted so they had little decrease in looks to the agent in favor of looks to the patient in the first two seconds analyzed. When considering only the older children, significantly different behavior of looking during the sentences with animate object was found, namely that the decrease in looking to the agent in favor of the patient from the first to the second time window was greater when the patient was animate.

The effect of animacy on gaze of the Dutch 3½-year-olds can be considered to be in line with the effect of animacy on the gaze of the Dutch adults. The Dutch adults switched their attention to the patient in the [+an -an] to the least degree compared to the other sentence types, an effect attributed to the speed of their [+an -an] trials. That is, the inanimate patients in [+an -an] sentences are "uninteresting," or the least demanding types of patients to process. The Dutch 3½-year-olds switched their attention to the patient to the greatest degree when the patient was animate, which might be attributed to the fact that animate patients are “interesting” and deserve more attention. We now turn to the production studies of the English-speaking participants.

## 7.4 Experiment 5b

### 7.4.1 Method

*Participants*

The same English-speaking adults ($n$ = 31) who participated in the picture selection task in Experiment 5a also participated in an elicitation task.

*Design and materials*

The same animations and design were used as in Experiment 3b. The same target sentences were elicited, but in English.

*Apparatus*

The computer set-up was the same as in Experiment 5a except that participants heard no audio, but rather produced sentences that were recorded on a microphone via E-Prime.

*Procedure*

The same testing procedure was used as in Experiment 3b.

*Scoring*

The same scoring procedure was employed as in previous “b” Experiments. An independent transcription of 10% of the participants ($n$ = 3) showed 88% agreement with the main transcription. Inter-scorer agreement was also high (Cohen’s $\kappa$ = .94).

### 7.4.2 Results

*Criteria for Inclusion*

There were 2 participants excluded because they did not contribute at least two scorable responses per sentence type, attributed to too many *Nontarget* responses. The removal of these participants decreased the number of participants in the analysis to 29.

*Accuracy*

Of the 464 responses given by English-speaking adults, 436 were scorable and 26 were unscorable (23 *Nontarget* and 3 *Unclear*), and 2 were *Passive*. Of the scorable responses, adults used *SO order* 100% of the time.

*Voice onset latency*

Items with extreme VOLs ($n$ = 3) were removed from the VOL analysis. Mean VOLs on the four sentence types based on participant means are shown in Table 7.5 and Figure 7.9. It took the adults 1470 ms on average to begin a sentence.

To determine whether S-O animacy affected speed of response in the elicited production task, the log transformed VOLs were fit to a model with subject and object animacy as fixed factors and participants and items as random factors. Since including an interaction was not justified ($\chi^2(1) = 2.13$, $p > .1$), the baseline model was checked for main effects. There was no significant main effect of either

**Table 7.5 Mean VOL for producing SO order: Experiment 5b (English-speaking adults)**

| Sentence type | Mean | *sd* |
|---|---|---|
| + animate + animate | 1513 | *517* |
| + animate - animate | 1397 | *417* |
| - animate + animate | 1449 | *337* |
| - animate - animate | 1537 | *517* |
| Total | 1470 | *396* |

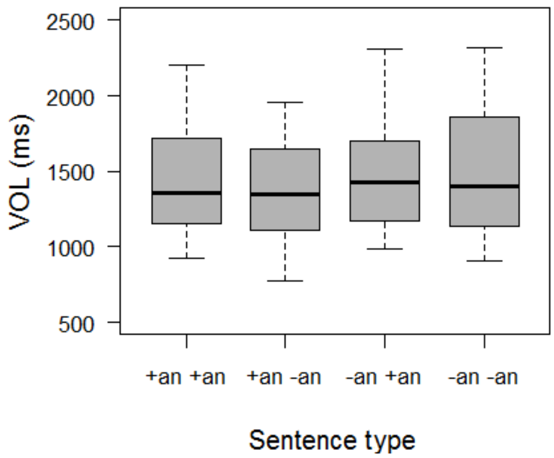


**Figure 7.9 Mean VOL for producing SO order: Experiment 6b (English-speaking adults)**

subject animacy ($\chi^2(1) = 1.80$, $p > .1$) or object animacy ($\chi^2(1) = 1.34$, $p > .1$). The inclusion of control factors such as gender, test verb, direction of action, and list did not significantly explain more variance in the data. Thus, the animacy manipulation did not affect how soon English adults began their sentences. The inclusion of first task, however, did significantly improve the baseline model ($\chi^2(1) = 15.35$, $p < .001$), and revealed that English-speaking adults had longer VOLs when they performed the elicitation task before the picture selection task ($\beta = .16$; $t = 4.29$; $p < .001$). In other words, they were able to begin their sentences faster in general when they were familiar with the animations and sentences.

*Eye movements*

Items that remained in the VOL analysis were used in the analysis of eye movements. There were 2 adults excluded because most of their trials had extreme track loss. Remaining items with extreme track loss ($n$ = 4) were also removed.

AOIs in the visual stimuli were defined over *Agent*, *Patient*, and *Not on AOI*. English-speaking adults showed the same general pattern of looking to the agent and then patient (shown in Figure 7.10) as the Dutch-speaking adults had. Two windows of time were defined for analysis: Time window 1 contains gaze data during 1000 ms prior to the onset of the sentence subject, and Time window 2 contains gaze data during 1000 ms after the onset of the sentence subject. For each participant and item, the agent advantage score was calculated. Figure 7.11 plots the mean agent advantage score in each time window. Negative scores occur in the second time window, indicating an advantage of the patient after the participants have begun their sentences.

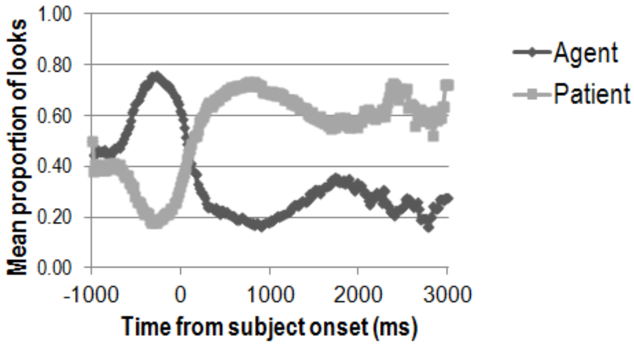


**Figure 7.10 Proportion of looks to agent and patient: Experiment 5b (English-speaking adults)**

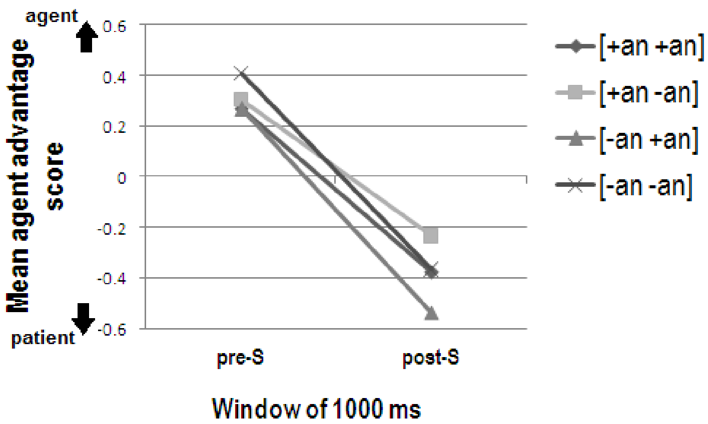


**Figure 7.11 Mean agent advantage score over two time windows: Experiment 5b (English-speaking adults)**

To determine whether S-O animacy affected which AOI was fixated during sentence planning and production, the empirical logit transformed mean agent advantage scores from each time window were fit to a model with subject animacy, object animacy, and time window as fixed factors, and with participant and item as random factors. There was no significant three-way interaction between the fixed predictors ($\chi^2(1) = .16$, $p > .1$), but there was a significant two-way interaction of time window and subject animacy ($\chi^2(1) = 5.87$, $p = .02$) which could be interpreted in light of a significant effect of time window ($\chi^2(1) = 238.06$, $p < .001$). In all sentence types, there was a decrease in the preference for agent over patient from the first to the second time window ($\beta = -4.61.$; $t = 16.70$; $p < .001$), but this decrease was less severe in sentences with animate subject ($\beta = -.67.$; $t = 2.42$; $p < .02$). Thus, English-speaking adults increased their looks to patient as their sentence unfolded, but to a lesser degree for the sentences with animate subjects. The inclusion of control factors such as test verb, first task, direction of action, and list showed that the inclusion of verb x time window significantly explained more variance in the data ($\chi^2(1) = 43.17$, $p < .001$), with greater looks to the agent over patient in the first time window when the verb was *pull* ($\beta = 3.09.$; $t = 6.46$; $p < .001$).

### 7.4.3 Discussion

Results of Experiment 5b show that English-speaking adults produced SO word order 100% of the time when describing the animations. Their VOLs, unlike those of Dutch adults, did *not* vary as a function of the manipulation of subject or object animacy. Like Dutch adults, they were faster overall to produce sentences when they received the production task after the comprehension task.

The gaze pattern of the English-speaking adults reflects the pattern found by Griffin and Bock (2000) for the production of simple, transitive sentences. For all sentence types, the adults looked first to the agent (the subject of the active sentence they produced) and began their sentence as they switched their gaze to the patient (the object of the active sentence they produced).The sentences with animate subjects saw the smallest decrease in looks to the agent in favor of the patient during one second after the sentence was started. Although no particular type of sentence was produced faster than another by the English-speaking adults, it could be the case that the less intense switch to patient in the [+an] subject sentences was due to the fact that these sentences required less attention to the patient for sentence completion. In other words, it could be that unusual agents or patients deserve more attention.

In sum, the English-speaking adults met the predictions that they would use SO order across the sentence types. They showed no facilitation effects of S-O animacy in their VOLs, but their gaze patterns can be interpreted as showing a facilitation effect of subject animacy. Thus, the inherent accessibility of animate entities resulted in quicker responses, but not word order alternations. We now turn to the production results from English-speaking preschoolers.

## 7.5 Experiment 6b

### 7.5.1 Method

*Participants*

The same English-speaking 2½-year-olds ($n$ = 19) and 3½-year-olds ($n$ = 12) who participated in the picture selection and preferential looking tasks in Experiment 6a also participated in an elicitation task.

*Design and materials*

The same target sentences (in English), animations, and design were used as in Experiment 4b.

*Apparatus*

The computer set up was the same as in Experiment 5b except that productions were recorded via a separate voice recorder instead of via E-Prime.

*Procedure*

The same testing procedure was employed as in Experiment 4b. The production procedure immediately followed the comprehension procedure described in Experiment 6a, which had been in turn preceded by a toy naming and verb demonstration pre-test. The only differences between this experiment and Experiment 4b is that the children were tested in Philadelphia at a day care, rather than in a lab in Groningen with a parent. Also, Penny the Pig stood in for Joris the Sheep as the eye-closing puppet.

*Scoring*

The same scoring procedure was employed as in previous “b” experiments. An independent transcription of 10% of the participants ($n$ = 3) showed 92% agreement with the main transcription. Inter-scorer agreement was also high (Cohen’s $\kappa$ = .86).

### 7.5.2 Results

*Criteria for inclusion*

Three children were tested in only one session because they were not present when the researchers returned for their second session; one of these children had to be removed because he did not contribute enough scorable responses from this single session. There were 5 additional children excluded because they did contribute at least two scorable responses per sentence type from their two sessions. The removal of these participants decreased the number of 2½-year-olds by 5 (*n* = 14, *mean* = 2;10, *sd* = 3.3 mo.s) and 3½-year-olds by 1 (*n* = 11, *mean* = 3;9, *sd* = 2.1 mo.s).

*Scorability*

The utterances produced by the remaining children were largely scorable. Of the 214 responses given by 2½-year-olds, 183 were scorable and 31 were unscorable*:* 21 *Nontarget*, 4 *Insufficient*, and 6 *Unclear*. (2 received *No answer* and 8 were *Not given*.) Of the 176 responses offered to 3½-year-olds, 152 were scorable and 8 were unscorable (all *Nontarget*).

*Accuracy*

Mean accuracy for each group on the four sentence types based on participant means are shown in Table 7.8 and Figure 7.12. On average, the 2½-year-olds used *SO order* 79% of the time (and *OS order* 21%), while the 3½-year-olds used *SO order* 94% of the time (and *OS order* 6%). (Individual scores can be found in Table C.15 and Table C.16. in Appendix C.)

**Table 7.8 Mean proportion SO order produced: Experiment 6b (English-speaking preschoolers)**

| Sentence type | 2½yrs | *sd* | 3½yrs | *sd* |
|---|---|---|---|---|
| + animate + animate | .82 | *.29* | .98 | *.08* |
| + animate - animate | .89 | *.17* | .91 | *.20* |
| - animate + animate | .57 | *.33* | .86 | *.26* |
| - animate - animate | .87 | *.28* | 1.00 | - |
| Total | .79 | *.18* | .94 | *.13* |

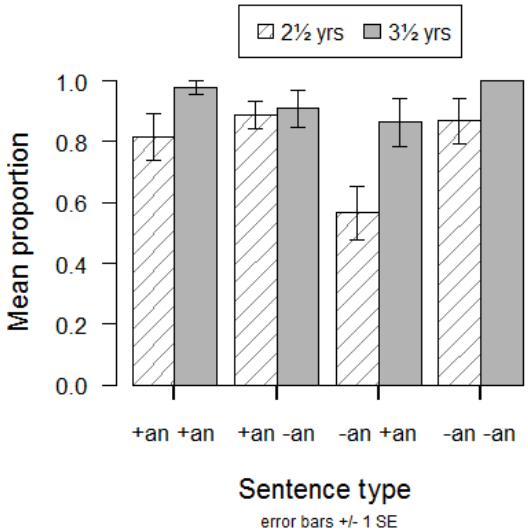


**Figure 7.12 Mean proportion SO order produced: Experiment 6b (English-speaking preschoolers)**

To determine whether S-O animacy affected accuracy in production in either age group, the binomial data were fit to a model with subject animacy, object animacy, and age group as fixed factors and participants and items as random factors. A fitting of the most complex model revealed that several predictors were completely collinear: subject animacy was a collinear predictor with an age group x subject animacy interaction, and object animacy was a collinear predictor with an age group x object interaction, and age group was a collinear predictor with an age group x subject animacy x object animacy interaction. In other words, the effect of age group was largely coming from the condition in which the subject was inanimate and the object was animate. Since the question of interest is whether the children are affected by animacy information in sentence production, all interaction terms with age group were removed as a strategy to reduce collinearity (Baayen, 2008: 183). Thus a model was fit with the three fixed predictors and only a single two-way interaction of subject animacy and object animacy.

There was a significant two-way interaction between subject animacy and object animacy ($\chi^2(1) = 4.31$, $p = .04$), which could be interpreted in light of a significant effect of object animacy ($\chi^2(1) = 6.77$, $p = .03$): all children were less likely to use SO order when the patient was animate ($\beta = -.49$; $z = -2.42$; $p = .02$), a negative effect that was greatly diminished when the agent was also animate ($\beta = .50$; $z = 2.49$; $p = .001$). Furthermore, there was a significant effect of age group ($\chi^2(1) = 6.29$, $p = .04$). This means that the older children were more likely to produce SO sentences than the younger children ($\beta = .33$; $t = 4.40$; $p < .001$). The inclusion of control factors such as gender, test verb, direction of action, and list showed that the inclusion of verb explained significantly more variance in the data ($\chi^2(1) = 4.30$, $p = .04$), with the children more likely to use SO order with the verb *pull* ($\beta = .40$; $z = 2.06$; $p = .04$).

*Eye movements*

All of the children included in the accuracy analysis were included in the production gaze analysis except for two children who had too many trials across conditions with extreme track loss. An item had extreme track loss if there was track loss of both eyes for more than one-thirds of three-second region of interest. Remaining items with extreme track loss ($n = 56$) and OS responses were also removed from the analysis. Each child contributed at least two responses per sentence type except for 10 children who contributed at least two responses on three sentence types and 3 children who contributed at least two responses on two sentence types.

AOIs in the visual stimuli were defined over *Agent*, *Patient*, and *Not on AOI*. Figure 7.13 and Figure 7.14 show the general pattern of looks to agent and patient over the course of a trial for each age group, synchronized to the onset of each child's sentence. The gaze plot shows that English preschoolers looked first to the agent (subject) prior to starting a sentence, and then to the patient (object). The crossover from agent to patient was protracted from subject onset by about 250 ms in both groups of children.

Two windows of time were defined for analysis: Time window 1 contains gaze data during 1000 ms prior to the onset of the sentence subject, and Time window 2 contains gaze data during 1000 ms after the onset of the sentence subject. For each participant and item, the agent advantage score was calculated. Figure 7.15 and 7.16 plots the mean agent advantage score in each time window for each age group. Negative scores indicating an advantage of the patient occur in the second time window, after the participants have begun their sentences.

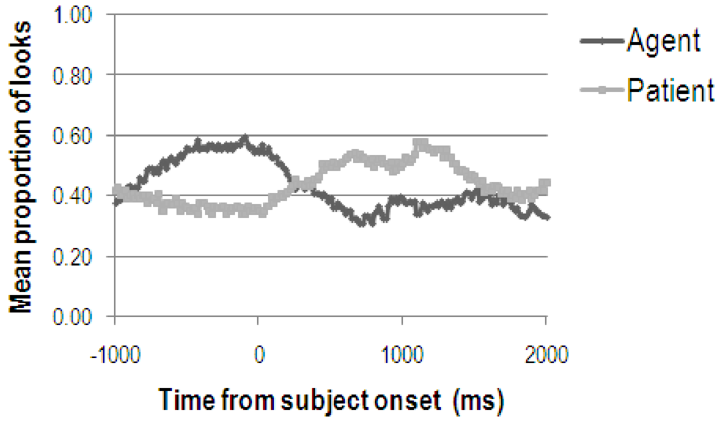


**Figure 7.13 Proportion of looks to agent and patient: Experiment 6b (English-speaking 2½ yrs)**

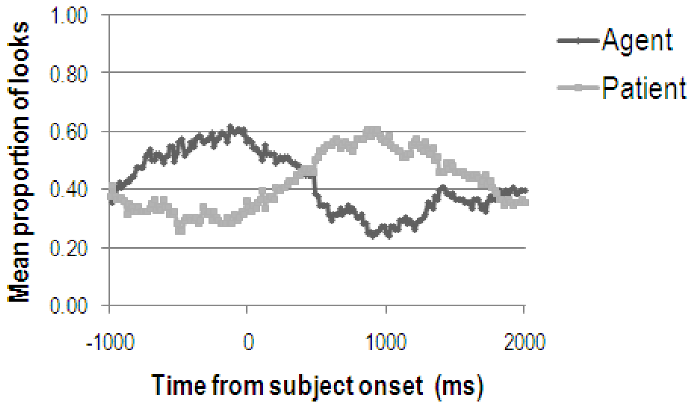


**Figure 7.14 Proportion of looks to agent and patient: Experiment 6b (English-speaking 3½ years)**

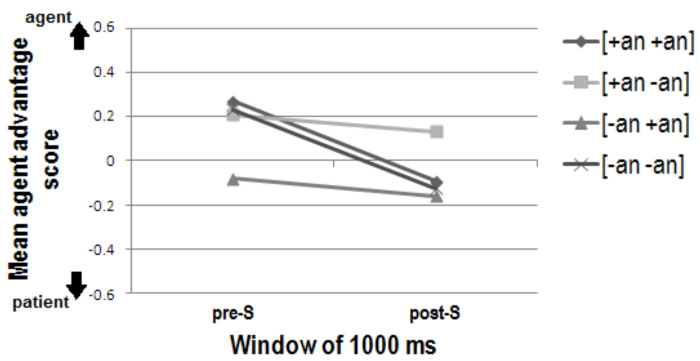


**Figure 7.15 Mean agent advantage score over two time windows: Experiment 6b (English-speaking 2½ yrs)**

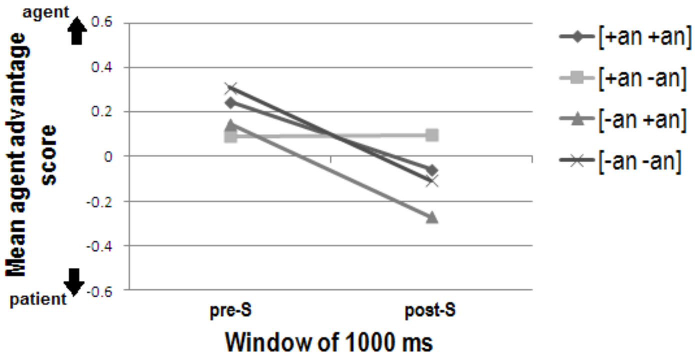


**Figure 7.16 Mean agent advantage score over two time windows: Experiment 4b (English-speaking 3½ yrs)**

To determine whether S-O animacy affected which AOI was fixated during sentence planning and production for each age group, the empirical logit transformed mean agent advantage scores from each time window were fit to a model with subject animacy, object animacy, time window, and age group as fixed factors, and with participant and item as random factors. There were no significant four-way ($\chi^2(1)$ = .08; $p > .1$), three-way ($\chi^2(4) = 5.03$; $p > .1$), or two-way ($\chi^2(6) = 3.31$; $p > .1$) interactions between the fixed predictors. In the baseline model, there was a significant main effect of time window ($\chi^2(1) = 12.36$; $p < .001$), with the agent advantage decreasing from the first to the second time window ($\beta = -1.57$; $t = -3.53$; $p < .001$). Furthermore, there was a significant main effect of subject animacy ($\chi^2(1) = 4.18$; $p = .04$) and object animacy ($\chi^2(1) = 4.60$; $p = .03$), indicating that the agent was more likely to be fixated in either time window when the subject was animate ($\beta = .59$; $t = 2.04$; $p = .04$) and when the object was inanimate ($\beta = .62$; $t = 2.13$; $p = .03$). The inclusion of control factors such as gender, test verb, direction of action, and list did not significantly explain more variance in the data.

### 7.5.3 Discussion

Results of Experiment 6b show that S-O word order used by English-speaking preschoolers, like the Dutch-speaking preschoolers, did vary as a function of the manipulation of subject and of object animacy. Basically, the combined advantageous effect of subject animacy with the detrimental effect of object animacy resulted in greater use of OS word order to describe S<O animations compared to the others. Such an effect of subject and object animacy was not predicted on the basis of Hendriks et al.'s (2005) OT model. However, on the whole, use of word order in production—certainly by the older children—was generally good: the 2½-year-olds produced SO word order 79% of the time 3½-year-olds 94% of the time

The gaze pattern of the English-speaking preschoolers was adult-like in that it reflected the pattern found by Griffin and Bock (2000) for simple, transitive sentences. For all sentence types, the children looked first to the agent (the subject of the active sentence they produced) and began their sentence as they switched their gaze to the patient (the object of the active sentence they produced). The children of both age groups began a sentence roughly 250 ms after they had viewed the agent. There were two distinct effects of animacy on this gaze pattern: one of subject animacy and one of object animacy. This essentially means that in both time windows, the children were more likely to look at the patient in the S<O animations. This is visualized in Figure 7.15 and Figure 7.16: the children looked less to the agent during the first time window for [-an +an] sentences relative to the other sentences, and looked more to the patient (signified by the lowest agent advantage score) during the second time window relative to the other sentences.

The effect of subject animacy and object animacy on gaze is an interesting reflection of the children's difficulty with using SO word order in the S<O sentence type, especially considering that only gaze from SO sentences were included in the gaze analysis. That is, even when the English-speaking preschoolers used SO word order to describe S<O animations, they were attending to the animate patient across the windows to a greater degree than they attended to any type of patient in the other types of animations. This unexpected result will be further addressed in the general discussion that now follows.

## 7.6 General Discussion

The questions that Experiments 3b – 6b aim to answer is (i) whether Dutch- and English-speaking preschoolers' word order is adult-like, (ii) whether preschoolers' and adults' word order is unaffected by the S-O animacy manipulation in the production task, and (iii) whether these four populations' expected agent-then-patient gaze pattern is affected by the S-O animacy manipulation. The results of experiments 3b – 6b, which tested sentences with an animal-vehicle distinction, are summed up as follows:

- Experiment 3b tested the production of 41 Dutch adults (from Experiment 3a) and Experiment 5b tested the production of 31 English-speaking adults (from Experiment 5a) using a sentence elicitation task with animations.
    - Adults of both languages produced SO word order 100% of the time.
    - Adults of both languages exhibited an agent-then-patient gaze pattern.
    - For Dutch adults, the S>O sentences saw the smallest decrease in looks to the agent in favor of the patient during one second after the sentence was started.
    - For English-speaking adults, the sentences with animate subjects saw the smallest decrease in looks to the agent in favor of the patient during one second after the sentence was started.

- Experiment 4b tested the production of 32 Dutch preschoolers (from Experiment 4a) using a sentence elicitation task with animations.
    - Both 2½- and 3½-year-old Dutch children produced SO word order over 80% of the time, but they were more likely to use SO order when the subject was animate than when it was inanimate.
    - Both 2½- and 3½-year-old Dutch children exhibited an agent-then-patient gaze pattern.
    - The older children began a sentence roughly 250 ms after they had viewed the agent, while the younger children began a sentence roughly 750 ms after they had viewed the agent.
    - For the older children, the decrease in looking to the agent in favor of the patient from the first to the second time window was greater when the patient was animate.

- Experiment 6b tested the production of 31 English-speaking preschoolers (from Experiment 6a) using a sentence elicitation task with animations.
    - Both 2½- and 3½-year-old English-speaking children produced SO word order over 79% of the time, but they exhibited a combined effect of subject and object animacy that caused the proportion SO word order produced in the S<O situations to suffer.
    - Both 2½- and 3½-year-old English-speaking children exhibited an agent-then-patient gaze pattern.
    - The children of both age groups began a sentence roughly 250 ms after they had viewed the agent.
    - The children of both age groups were more likely to look at the patient in the S<O animations

In the following discussion, these results are interpreted in light of the questions asked. First, word order responses are considered, and then effects of animacy on gaze are addressed.

*Effect of animacy on word order production*

Is Dutch- and English-speaking preschoolers' word order adult-like? Hendriks et al.'s (2005) model of S-O word order development predicts SO word order for preschool speakers of Dutch and English. Preschoolers of both age groups used SO word order over 79% of the time, indicating overall adherence to the word order rules of the target language.

Is preschoolers'and adults' word order unaffected by the S-O animacy manipulation in the production task? Hendriks et al.'s model predicts SO word order across the four sentence types for adult and preschool speakers of Dutch and English. For both Dutch- and English-speaking preschoolers, however, there was a significant effect of animacy on the word order they produced. The Dutch preschoolers were less likely to use SO order when the agent was a vehicle than when it was an animal; the English children were least likely to use SO order to describe animations in which the agent was a vehicle and the patient was an animal. These effects of animacy on word order used in production were not predicted and will be addressed further in the discussion in Chapter 9.

The unexpected effect of animacy in the word order used by preschoolers of Dutch and English does not preclude the possibility of an asymmetry between comprehension and production. The Dutch preschoolers in Experiment 4b produced SO word order over 80% of the time on average across sentence types. This contrasts with the data from the picture selection task with the same population presented in Experiment 4a, in which there was a distinct effect of both subject animacy and object animacy, and the average number of SO interpretations on any sentence type did not reach 80% in the picture selection task. Likewise, the effect of S-O animacy found in the word order produced by English-speaking children in Experiment 6b contrasts with the data from the picture selection task with the same population tested in Experiment 6a, in which there was no effect of animacy on word order interpretation—the inverse of the pattern expected. The differences between the proportion of SO order used in comprehension versus production in these populations will be analyzed in Chapter 8.

If we compare the elicitation task used with animations to the elicitation task used with toys in the act-out tasks described in Chapter 5, we notice that roughly the same number of children were excluded from the production analyses. Experiments 1b and 2b with Dutch children saw about one third of the children excluded due to too many unscorable responses or non-responses. The same held for the Dutch children tested in Experiment 4b. The English children fared a bit better, with a loss of roughly one-fifth of the children in Experiment 6b. Another notable difference was that the children in the act-out study did not seem to be affected by the control factors, whereas there were some inconsistent influences of control factors in the animation description task, namely that the Dutch children used SO word order more often when the action was pushing, and when the direction of action was to the left; the English-speaking children used SO word order more often when the action was pulling. There is no explanation for why the children would be influenced by the control factors in these inconsistent ways, but since these factors were controlled across the different sentence types, the effects are not seen as interfering with the interpretation of the overall results.

Turning to adults: predictions were met since SO order was overwhelmingly used across sentence types. There were only two passive sentences produced in either language, a figure that is strikingly low considering the results of previous studies showing that adults tend to place animate patients early in the sentence either via passive structures or object-fronting (e.g. Ferreira, 1994; Griffin & Bock, 2000; Hartsuiker et al., 2004). Unlike such studies that compare active and passive sentences, the current study did not explicitly encourage the production of passives by including structural priming of passives, making certain entities larger and therefore more salient, or using verbs with the requirement of either an animate or inanimate agent. Thus, when encouraged to produce simple, active sentences with the reversible verbs *push* and *pull*, speakers of English and Dutch do not prefer animals to be earlier in the sentence than vehicles. Thus the difference in inherent accessibility of animals versus vehicles does not trigger structural alternations when there is no further contextual, discourse-level reason to do so.

*Effect of animacy on gaze during production*

Griffin and Bock (2000) found that when there is no linguistic task, participants simply look at depicted events in a scene, but when there is a sentence production task, an agent-then-patient gaze pattern arises. The pattern of gaze exhibited by all participants in the animation description tasks in Experiments 3b – 6b corroborates Griffin and Bock's finding that gaze is linguistically mediated: speakers first looked to the agent, and then looked to the patient as they began a sentence. This may be the first study to show that preschooler-aged speakers of English and Dutch show the same pattern as adults, albeit with some protraction of the switch from agent to patient in relation to their VOL. Thus, in the task where an animation had to be described, neither the adults nor the children in the present studies randomly scanned the scenes, but rather they sought out the entity that would be their active sentence's subject and then the entity that would be object. This may explain why all of the participants, save the Dutch preschoolers, had the strongest looks to pulling agents in the one second prior to sentence production, since the rope involved in the pulling action may have made the pulling agent easier to identify.

Was this agent-then-patient gaze pattern affected by the S-O animacy manipulation? Animations with animate agents were expected to be facilitated during production due to the inherent accessibility of animate entities. There were indeed some facilitative effects of S-O animacy on the agent-then-patient pattern found in the different populations. First, let us look at the populations that used SO order across sentence types (adults of both languages and Dutch preschoolers). There was a general tendency for the adults to show a decrease in looks to the agent in favor of looks to the patient, a switch that was the least intense in sentences that had (i) prototypical agents and patients ([+an -an] for Dutch adults), or (ii) prototypical agents (animate agents for English-speaking adults). It was suggested that sentences are easy to produce because of the accessibility of animate entities, so the attention to the patient need not be great. To the same end, the Dutch 3½-year-olds showed a decrease in looks to the agent in favor of looks to the patient that was the most intense when the patient was not prototypical (animations with animate patients) and perhaps required more attention.

The English-speaking preschoolers instead actually used (ungrammatical) OS word order frequently to describe S<O animations, which is corroborated by the fact that they attended to animate patients in these types of animations to a greater degree than they attended to any type of patient in the other types of animations. These results are unexpected and will be addressed further in the discussion in Chapter 9.

The effect of animacy on gaze during production differs from the effect of animacy on within-animation gaze when viewing animation pairs during comprehension. In the production task, the agent-then-patient pattern was inhibited or facilitated by S-O prototypicality, but participants still sought the agent first and then the patient regardless of their animacy. In the comprehension task (picture selection for adults and preferential looking for children), participants tended to look more at the most alive entity within each animation in their search for agent. Remember that they showed a preference for agents in S=O animation pairs, but a preference for animate entities in S≠O animation pairs. Thus, in production there is a preference to find agents in order to begin an utterance, while in comprehension there is an expectation that animate entities are agents. The implication of these findings are discussed in Chapter 9.

In sum, adult speakers of Dutch and English adhered to SO word order across sentence types while preschoolers were unexpectedly affected by the animacy manipulation in their production of word order. The gaze data from all populations corroborated the agent-then-patient pattern during active sentence production found by Griffin and Bock (2000), a pattern that for all populations but the English-preschoolers was facilitated by the inherent accessibility of animate entities. The animacy effects on gaze during production was interpreted as different in nature from the animacy effects on within-animation gaze during comprehension. Now that we have looked at six sets of comprehension and production experiments, we are in a position to compare the results from the two types of tasks carried out by each population in the next chapter.

# 8 Comprehension vs. production

A comparison of Experiments 1 – 6

So far, the results of the comprehension experiments and the production experiments in Chapters 4 – 7 have been interpreted individually. In this chapter, the results from the a-experiments that tested comprehension are compared to the b-experiments that tested production with regard to scorability and accuracy. (An overview of all experiments is provided in Table 1.1 at the end of Chapter 1.)

## 8.1 Introduction

Each experiment in the previous chapters was already inspected for an effect of subject or object animacy. Now performance on the comprehension and production tasks in general will be compared in order to determine if performance on one exceeded performance on the other. Better performance on either production or comprehension, especially across different paradigms, would be seen as evidence for asymmetry.

If it is to be argued that performance on production is better than comprehension, it is important to address the fact that the production task yielded less data than the comprehension task. In the four studies with preschoolers presented in Chapters 4 – 7, less children were able to be included in the production analyses than the comprehension analyses due to too many unscorable or missing responses. In the first part of this chapter, scorability of each task is inspected, for both adults and children. If children gave more scorable responses on the comprehension task than the production task, this can be seen as evidence that the production task is more open-ended, and perhaps not as easy to perform as the comprehension task.

Once the scorability of the tasks has been compared, the accuracy on each of the tasks is then compared. In an item-based analysis, the accuracy on items (particular sentences) that received a scorable response on both comprehension and production by a participant will be looked at. If children produced SO word order more often than they understood word order as SO with this set of items, this will be seen as evidence that word order production exceeds and precedes that of comprehension by the children in these studies.

## 8.2 Scorability

There appeared to be many unscorable responses in the production tasks in comparison to the comprehension tasks, which was not unexpected, as the scorability per task calculated from the data of C&M also showed that children responded with more scorable responses in an act-out task than in a corresponding elicitation task (See section Figure 3.1 in Section 3.3.3). One could argue, on the basis of this sheer difference in scorability on the two tasks, that production is in fact not as good as comprehension in preschoolers. However, production tasks are by nature more free than comprehension tasks in that the possibilities of what will be uttered are limitless. This problem is intrinsic to studies that elicit production.

If the comprehension tasks yielded more scorable responses than the corresponding production tasks, this goes against an extra-grammatical account of asymmetries that appeals to experimental limitations. That is, if better production than comprehension of word order is seen as a result of one task being easier than the other, one would expect that production was in fact the easier of the two tasks. However, many younger children were able to carry out a comprehension task, but produced insufficient utterances during the elicitation task. Thus, higher scorability on a comprehension task is at odds with the argument that an apparent asymmetry (here, S-O word order) results from one task simply being easier (here, production).

In order to confirm that scorability on comprehension was higher than in production, the tasks are compared within each experiment. First, scorability in each experiment is checked for an effect of task or age group. Then scorability of each task in each experiment is analyzed individually to see whether there was an influence of age group or S-O animacy. It would be natural for older children to give more scorable responses than younger children in general, at least on the production tasks which usually saw a decrease in the set of 2½-year-olds included in the analyses. If children gave more scorable responses on the comprehension task than the production task, it can be argued that production is not an easier task than comprehension.

In the following analyses, scorability is broken down into the broad categories of scorable, unscorable, or missing. *Scorable* means SO/OS interpretations in comprehension or SO/OS order in production; *Unscorable* means unscorable actions/picture selections in comprehension or unscorable utterances in production; and *Missing* means items were not answered or administered.

*Adults*

A breakdown of the types of response scorability in each task performed by adults in Experiments 3 and 5 is presented in Table 8.1. These numbers are based on all adults included in the comprehension analysis[18], even if some were to be excluded in the production analysis. While there were no unscorable responses in comprehension (since these were given by button-press in the picture selection task), there were utterances that were unscorable in the elicited production task. Table 8.2 breaks down scorability during the production task by sentence type. To determine whether S-O animacy affected scorability in production, the binomial data (scorable vs. unscorable) were fit to a mixed effects model for each group of adults. Missing data were not taken into account. Subject animacy and object animacy were included as fixed factors and participants and items as random factors.

In Experiment 3b, an interaction of subject animacy and object animacy approached significance ($\chi^2(1) = 3.57$, $p = .06$). In the baseline model, there was no main effect of object animacy ($\chi^2(1) = 1.08$, $p > .1$), but there was a significant effect of subject animacy ($\chi^2(1) = 5.84$, $p = .02$), with Dutch adults more likely to produce a scorable response when the subject was animate ($\beta = .76$; $z = 2.37$; $p = .02$). The inclusion of control factors showed that first task significantly explains more variation in the data ($\chi^2(1) =$

[18] The starting point for a scorability analysis with a particular population is the group of participants who were included in a comprehension analysis. For the adult populations, no adult was excluded from a comprehension analysis, so the starting point is all adults tested. For child populations, a few children (never more than two per population) were excluded from each comprehension analysis, so the starting point for the scorability analysis is slightly less than all children tested. This was done because (i) the few excluded children were always also excluded from the accompanying production analysis, so did not contribute to any accuracy results reported, and (ii) the scorability analyses is then run with the same set of children with which accuracy analyses have been run on in previous chapters.

17.97, $p < .001$). Dutch adults were less likely to produce a sentence with scorable word order when they received the production task first ($\beta = -1.31$; $z = -3.93$; $p < .001$).

**Table 8.2 Mean proportion of scorable responses in production per condition: Experiments 3 & 5**

| Experiment | Population | Production | | | |
|---|---|---|---|---|---|
| | | [+an +an] | [+an -an] | [-an +an] | [-an -an] |
| 3b | Dutch-speaking adults | .93 | .99 | .91 | .88 |
| | | *(.17)* | *(.08)* | *(.20)* | *(.17)* |
| 5b | English-speaking adults | .91 | .99 | .90 | .86 |
| | | *(.24)* | *(.04)* | *(.15)* | *(.17)* |

In Experiment 5b, there was a significant interaction of subject animacy and object animacy ($\chi^2(1) = 6.19$, $p = .01$) as well as an effect of subject animacy ($\chi^2(1) = 6.17$, $p = .01$). This means that English-speaking adults were more likely to produce a scorable response when the subject was animate ($\beta = .87$; $z = 2.08$; $p = .04$), an effect that was diminished somewhat in [+an +an] sentences where the object was also animate ($\beta = -.86$; $z = 2.06$; $p = .04$). The inclusion of control factors showed that first task ($\chi^2(1) = 12.80$, $p < .001$) and verb ($\chi^2(1) = 6.67$, $p = .01$) significantly explain more variation in the data. English-speaking adults were less likely to produce a sentence with scorable word order when they received the production task first ($\beta = -1.13$; $z = -3.48$; $p < .001$), and they were more likely to produce a sentence with scorable word order when describing *pull* actions ($\beta = .75$; $z = 2.68$; $p = .008$).

In sum, there was no variation in the scorability of the comprehension responses given by adult speakers of Dutch and English. Production responses, on the other hand, were unscorable between 6% and 8% of the time. A closer look at the unscorable utterances showed that Dutch adults were more likely to describe animations with inanimate agents (versus animate agents) with an unscorable utterance and that English-speaking adults were most likely to produce unscorable utterances when describing S<O animations. Furthermore, familiarization with the animations during a comprehension task received before a production task increased the likelihood that an animation would be described with an utterance that contained scorable word order.

*Preschoolers*

A breakdown of the types of response scorability in each task performed by children in Experiments 1, 2, 4, and 6 is presented in Table 8.3. Responses in comprehension refer to actions in Experiments 1a and 2a and to points in Experiments 4a and 6a. Responses in production refer to the description of actions with toys in Experiments 1b and 2b and to the description of animations in Experiments 4b and 6b. Average scorability on the comprehension tasks ranged between 92% and 98%, whereas average scorability on the production tasks ranged between 30% and 88%. These numbers are based on the children who were included in each comprehension analysis, even if they were to be excluded in the production analysis.

In order to determine if there was an effect of task (comprehension vs. production) on scorability in either age group, the binomial data (scorable vs. unscorable) were fit to a model with scorability and age group as fixed factors and items as random factors. Missing data was not taken into account. The coefficients of the full model for each experiment are shown in Table 8.4. The interaction of age group and task was excluded from the model for English-speaking children because there was complete collinearity between the interaction and the individual factors. Remember that a $\chi^2$-value is reported to show whether the inclusion of a factor adds to the explanation of variance in a model, and a $\beta$-value of a

**Table 8.1 Number of scorable responses in comprehension and production tasks in Experiments 3 & 5**

| Experiment | Population | *n* | Trials | Comprehension (a) | | | Production (b) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Scorable | Unscorable | Missing | Scorable | Unscorable | Passive | Missing |
| 3 | Dutch adults | 41 | 656 | 655 (99.8%) | 0 (0%) | 1 (0.2%) | 608 (92.7%) | 42 (6.4%) | 6 (0.9%) | 0 (0.0%) |
| 5 | English adults | 31 | 496 | 495 (99.8%) | 0 (0%) | 1 (0.2%) | 455 (91.7%) | 39 (7.9%) | 2 (0.4%) | 0 (0.0%) |

**Table 8.3 Number of scorable responses in comprehension and production tasks: Experiments 1, 2, 4 & 6**

| Experiment[1] | Population | *n* | Trials[2] | Comprehension (a) | | | Production (b) | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Scorable | Unscorable | Missing | Scorable | Unscorable | Missing |
| 1 (toys) | Dutch 2½ yrs | 12 | 192 | 177 (92.2%) | 14 (7.3%) | 1 (0.5%) | 69 (36.0%) | 92 (48.0%) | 31 (16.0%) |
| | Dutch 3½ yrs | 16 | 256 | 252 (98.4%) | 3 (1.2%) | 1 (0.4%) | 216 (84.4%) | 31 (12.1%) | 9 (3.5%) |
| 2 (toys) | Dutch 2½ yrs | 30 | 480 | 444 (92.5%) | 36 (7.5%) | 0 (0%) | 146 (30.4%) | 226 (47.1%) | 108 (22.5%) |
| | Dutch 3½ yrs | 33 | 528 | 511 (96.8%) | 14 (2.7%) | 3 (0.5%) | 388 (73.5%) | 134 (25.4%) | 6 (1.1%) |
| 4 (animations) | Dutch 2½ yrs | 13 | 208 | 195 (93.8%) | 10 (4.8%) | 3 (0.4%) | 96 (46.2%) | 78 (37.5%) | 34 (16.3%) |
| | Dutch 3½ yrs | 17 | 272 | 266 (97.8%) | 4 (1.5%) | 2 (0.7%) | 239 (87.9%) | 25 (9.2%) | 8 (2.9%) |
| 6 (animations) | English 2½ yrs | 18 | 280 | 267 (95.4%) | 2 (0.7%) | 11 (3.9%) | 188 (67.1%) | 65 (23.2%) | 27 (9.7%) |
| | English 3½ yrs | 12 | 176 | 171 (97.2%) | 0 (0%) | 5 (2.8%) | 136 (77.3%) | 27 (15.3%) | 13 (7.4%) |

1. Experiments 1a and 2b reflect responses in an act-out task with toys and Experiments 4a and 6a reflect responses in a picture selection task with animations. Experiments 1b and 2b reflect responses in a sentence elicitation task with toys and Experiments 4b and 6b reflect responses in a sentence elicitation task with animations.
2. The number of trials in English does not correspond to 16 trials per child because 3 children were tested in only one session.

significant factor is reported to show the direction of that factor's effect in the final model. The *p*-values of the two statistics for a factor are usually the same or very close.

**Table 8.4 Effects of task and age group on scorability: Experiments 1, 2, 4 & 6**

| Experiment | Factor | $\chi^2(1)$ | | $\beta$ | | |
|---|---|---|---|---|---|---|
| 1 (toys) | task * age group | .14 | $p > .1$ | .08 | $z = .37$ | $p > .1$ |
| | task | 187.81 | $p < .001$ | -1.96 | $z = -8.82$ | $p < .001$ |
| | age group | 12.80 | $p = .002$ | 2.31 | $z = 3.46$ | $p < .001$ |
| 2 (toys) | task* age group | 10.93 | $p < .001$ | .44 | $z = 3.25$ | $p = .001$ |
| | task | 582.94 | $p < .001$ | -2.23 | $z = -16.34$ | $p < .001$ |
| | age group | 42.40 | $p < .001$ | 1.18 | $z = 3.47$ | $p < .001$ |
| 4 (animations) | task* age group | 3.41 | $p = .06$ | .35 | $z = 1.92$ | $p = .06$ |
| | task | 130.99 | $p < .001$ | -1.41 | $z = -7.79$ | $p < .001$ |
| | age group | 19.42 | $p < .001$ | 1.19 | $z = 3.28$ | $p = .001$ |
| 6 (animations) | task* age group | - | | - | | |
| | task | 173.50 | $p < .001$ | -2.77 | $z = -7.27$ | $p < .001$ |
| | age group | .70 | $p > .1$ | | | |

Each experiment showed task to be a highly significant factor, with children less likely to give scorable answers on the production task than the comprehension task. This can be seen from the negative coefficients signaling a decrease in scorability for the production task. In Experiment 6 carried out with English-speaking children, there was no additional effect of age group. In Experiments 1, 2, & 4 carried out with Dutch children, there was an effect of age group, with the older children more likely to give scorable answers on the two tasks than the younger. In addition, the negative effect of the production task proved to be less severe for the older group of children than the younger group in Experiments 2 and 4. This can be seen from the positive coefficients of the task x age group interaction. Thus, across the different paradigms, responses in comprehension were more likely to be scorable than in production. Furthermore, Dutch 3½-year-olds were more likely to give scorable responses than their 2½-year-old peers, especially in the production task.

A closer look was taken at scorability on the different sentence types in the comprehension tasks. Table 8.5 shows the mean scorability per sentence type based on participant means in each comprehension experiment. These numbers are still based on the children who were included in the accuracy analysis for each comprehension task. Again, we see that average scorability on the comprehension tasks ranges between 90% and 100%.

To determine whether S-O animacy affected scorability on the comprehension tasks performed by either age group, the binomial data (scorable vs. unscorable) were fit to a mixed effects model for each a-experiment. Subject animacy, object animacy, and age group were included as fixed factors and participants and items as random factors. The responses of the older English-speaking children were excluded because there was zero variability in the scorability of their responses.

The coefficients for the four models are shown in Table 8.6. There were no significant three-way or two-way interactions between the fixed predictors, so only baseline results are listed. There was an effect of age group that approached significance in the picture selection task carried out with Dutch-speaking children, indicating that older children tended to point in a more scorable fashion than the younger children. But on the whole, neither age group nor the animacy of subject or object had a significant influence on the models of scorability in the different comprehension tasks.

**Table 8.5 Mean proportion of scorable responses in comprehension per condition: Experiments 1a, 2a, 4a & 6a**

| Experiment[1] | Population | n | Comprehension [+an +an] | [+an -an] | [-an +an] | [-an -an] |
|---|---|---|---|---|---|---|
| 1a | Dutch 2½ yrs | 12 | .96 | .96 | .96 | .93 |
| | | | *(.10)* | *(.10)* | *(.10)* | *(.17)* |
| | Dutch 3½ yrs | 16 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | | - | - | - | - |
| 2a | Dutch 2½ yrs | 30 | .92 | .94 | .92 | .95 |
| | | | *(.18)* | *(.16)* | *(.15)* | *(.14)* |
| | Dutch 3½ yrs | 33 | .97 | .98 | .96 | .98 |
| | | | *(.10)* | *(.09)* | *(.13)* | *(.10)* |
| 4a | Dutch 2½ yrs | 13 | .92 | .96 | .96 | .96 |
| | | | *(.16)* | *(.09)* | *(.09)* | *(.09)* |
| | Dutch 3½ yrs | 17 | .99 | .97 | 1.00 | .99 |
| | | | *(.06)* | *(.12)* | - | *(.06)* |
| 6a | English 2½ yrs | 18 | .97 | .98 | .97 | 1.00 |
| | | | *(.08)* | *(.08)* | *(.11)* | - |
| | English 3½ yrs | 12 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | | - | - | - | - |

1. Experiments 1a and 2b reflect responses in an act-out task with toys and Experiments 4a and 6a reflect responses in a picture selection task with animations.

**Table 8.6 Effects of age group and animacy on scorability in comprehension: Experiments 1a, 2a, 4a & 6a**

| Experiment[1] | Factor | $\chi^2(1)$ | | $\beta$ | | |
|---|---|---|---|---|---|---|
| 1a | age group | 1.10 | $p > .1$ | | | |
| | subject animacy | .16 | $p > .1$ | | | |
| | object animacy | .16 | $p > .1$ | | | |
| 2a | age group | 2.77 | $p = .1$ | | | |
| | subject animacy | .50 | $p > .1$ | | | |
| | object animacy | 1.11 | $p > .1$ | | | |
| 4a | age group | 3.53 | $p = .06$ | .73 | $z = 1.69$ | $p = .09$ |
| | subject animacy | 1.23 | $p > .1$ | | | |
| | object animacy | 0 | $p > .1$ | | | |
| 6a | age group | - | | | | |
| | subject animacy | .12 | $p > .1$ | | | |
| | object animacy | 0 | $p > .1$ | | | |

1. Experiments 1a and 2a reflect responses in an act-out task with toys and Experiments 4a and 6a reflect responses in a picture selection task with animations.

A closer look was taken at scorability on the different sentence types in the production tasks. Since we are interested in potential variation due to sentence type on children's ability to produce scorable responses, data is used from the set of children who were included in each experiment's production accuracy analysis since these were the children who did not have a severe number of missing or unscorable responses. Table 8.7 shows the mean scorability per sentence type based on participant means in each production experiment. As was the case in the larger set of children, the mean scorability on the production tasks remains quite variable in this subset, ranging from between 72% to 100% on individual sentence types.

**Table 8.7 Mean proportion of scorable responses in production per condition: Experiments 1b, 2b, 4b & 6b**

| Experiment[1] | Population | n | Production | | | |
|---|---|---|---|---|---|---|
| | | | [+an +an] | [+an -an] | [-an +an] | [-an -an] |
| 1b | Dutch 2½ yrs | 5 | .90 | .85 | .85 | .80 |
| | | | *(.22)* | *(.22)* | *(.22)* | *(.27)* |
| | Dutch 3½ yrs | 13 | 1.00 | 1.00 | .98 | 1.00 |
| | | | - | - | *(.07)* | - |
| 2b | Dutch 2½ yrs | 10 | .72 | .83 | .78 | .75 |
| | | | *(.22)* | *(.23)* | *(.22)* | *(.20)* |
| | Dutch 3½ yrs | 23 | .85 | .95 | .95 | .91 |
| | | | *(.21)* | *(.15)* | *(.15)* | *(.14)* |
| 4b | Dutch 2½ yrs | 5 | .85 | .95 | .90 | .80 |
| | | | *(.22)* | *(.11)* | *(.14)* | *(.11)* |
| | Dutch 3½ yrs | 16 | .92 | .92 | .88 | .91 |
| | | | *(.15)* | *(.17)* | *(.20)* | *(.18)* |
| 6b | English 2½ yrs | 14 | .84 | .91 | .87 | .80 |
| | | | *(.19)* | *(.19)* | *(.17)* | *(.22)* |
| | English 3½ yrs | 11 | .77 | .75 | .81 | .75 |
| | | | *(.39)* | *(.40)* | *(.40)* | *(.40)* |

1. Experiments 1b and 2b reflect responses in a sentence elicitation task with toys and Experiments 4b and 6b reflect responses in a sentence elicitation task with animations.

To determine whether S-O animacy affected scorability on the production tasks performed by either age group, the binomial data (scorable vs. unscorable) were fit to a mixed effects model for each b-experiment. Subject animacy, object animacy, and age group were included as fixed factors and participants and items as random factors.

The coefficients of the four models are shown in Table 8.8. Any three-way or two-way interactions between the fixed predictors that were significant or approached significance were included. Age group had an effect in the two experiments that elicited production via an action description task, with the older children more likely to produce a scorable utterance when describing actions depicted with toys than the younger. There was no such effect of age group found for the animation description task for either the Dutch- or English-speaking group. Regarding animacy, there was a significant interaction of subject animacy and object animacy in the second act-out task, with [+an +an] actions least likely to receive scorable descriptions. The same effect of S-O animacy only approached significance for the English-speaking children in the picture selection task. Overall, there was little effect of subject or object animacy on scorability in production.

In sum, children were more likely to give scorable responses in comprehension than in production, across languages and experimental paradigms. Age group sometimes played a role in scorability, usually among the Dutch populations. Overall, neither animacy of subject nor animacy of object significantly affected scorability in the different comprehension or production tasks. Now that it has been established that there is a difference between comprehension and production with regard to scorability, I look next at the difference between comprehension and production with regard to accuracy. Since there are more scorable responses in comprehension than production, the analysis is based on the set of items for which there was a scorable response on both comprehension and production.

**Table 8.8 Effect of age group and animacy on scorability in production: Experiments 1b, 2b, 4b & 6b**

| Experiment[1] | Factor | $\chi^2(1)$ | | $\beta$ | | |
|---|---|---|---|---|---|---|
| 1b | age group | 5.63 | $p = .02$ | 1.96 | $z = 2.17$ | $p = .03$ |
| | subject animacy | 1.06 | $p > .1$ | | | |
| | object animacy | .10 | $p > .1$ | | | |
| 2b | age group | 5.09 | $p = .02$ | .79 | $z = 2.41$ | $p = .02$ |
| | subject animacy | .47 | $p > .1$ | | | |
| | object animacy | 1.47 | $p > .1$ | | | |
| | subject an. * object an. | 3.92 | $p = .05$ | -.37 | $z = 2.12$ | $p = .03$ |
| 4b | age group | 1.53 | $p > .1$ | | | |
| | subject animacy | .17 | $p > .1$ | | | |
| | object animacy | 1.41 | $p > .1$ | | | |
| 6b | age group | .40 | $p > .1$ | | | |
| | subject animacy | 1.69 | $p > .1$ | | | |
| | object animacy | 1.48 | $p > .1$ | | | |
| | subject an. * object an. | 3.08 | $p = .08$ | -.37 | $z = -1.79$ | $p = .07$ |

1. Experiments 1b and 2b reflect responses in a sentence elicitation task with toys and Experiments 4b and 6b reflect responses in a sentence elicitation task with animations.

## 8.3 Item-based comparison

Last, but not least, the following analyses get to the root of the issue of this dissertation. Do children use SO order in production more successfully than they have SO interpretations in comprehension? The analyses include accuracy scores for items (particular sentences) for which a scorable response was obtained from a participant for both comprehension and production. By doing this, we can see how production compares with comprehension based on items that participants experienced in both tasks.

Animacy is not included as a factor in this final analysis. The effect of subject and object animacy on comprehension (Chapters 4 and 6) and on production (Chapters 5 and 7) was inspected in previous analyses. Now we are interested in the overall effect of task (comprehension versus production) on word order.

*Adults*

Items for which there was a scorable response for both comprehension and production (Dutch: $n$ = 608; English: $n$ = 455) were inspected for all adults in Experiment 3 and 5. Table 8.9 shows how these items were answered by each group of adults. On these items, there were few OS responses (between 0% and 3%) given in the two tasks.

**Table 8.9 Items with scorable response for comprehension and production: Experiments 3 & 5**

| | | Comprehension | Production |
|---|---|---|---|
| Dutch-speaking adults | SO | 589 (97.0%) | 607 (100%) |
| | OS | 18 (3.0%) | 0 (0%) |
| English-speaking adults | SO | 446 (97.4%) | 453 (99.8%) |
| | OS | 8 (2.6%) | 1 (0.2%) |

*Preschoolers*

Items for which there was a scorable response for both comprehension and production by a preschooler were analyzed. For each of the four experiments carried out with preschoolers, a breakdown of response type (SO vs. OS) by task is given for these items. Any child who was excluded from both the comprehension and production analyses of an experiment was not included in this item-based analysis.

In Experiment 1, there were 65 items for 2½-year-olds and 216 for 3½-year-olds for which there was a scorable response for both comprehension and production. Table 8.10 shows how these items were answered by each age group. In order to determine if there was a difference in word order responses (SO vs. OS) between the tasks on the basis of these items, the binomial data were fit to a model with task and age group as fixed factors and participant and item as random factors. There was no significant interaction of task and age group $\chi^2(1) = 1.03$, $p > .1$), so the baseline model was checked for main effects. There was a significant effect of task $\chi^2(1) = 15.18$, $p < .001$), with children more likely to give SO responses in production than comprehension ($\beta = .69$; $z = 3.66$; $p < .001$). There was also a significant effect of age group ($\chi^2(1) = 28.79$, $p$ <001), with older children more likely to give SO responses than younger children ($\beta = 1.46$; $z = 7.60$; $p < .001$).

**Table 8.10 Items with scorable response for comprehension and production: Experiment 1 (toys)**

| | | Comprehension | Production |
|---|---|---|---|
| 2½ yrs Dutch | SO | 39 (60.0%) | 54 (83.1%) |
| | OS | 26 (40.0%) | 11 (16.9%) |
| 3½ yrs Dutch | SO | 204 (95.8%) | 212 (99.5%) |
| | OS | 9 (4.2%) | 1 (0.5%) |

In Experiment 2, there were 143 items for 2½-year-olds and 383 for 3½-year-olds for which there was a scorable response for both comprehension and production. Table 8.11 shows how these items were answered by each age group. There was no significant interaction of task and age group $\chi^2(1) = .21$, $p > .1$). In the baseline model, there was a significant effect of task $\chi^2(1) = 22.84$, $p < .001$), with children more likely to give SO responses in production than comprehension ($\beta = .49$; $z = 4.67$; $p < .001$). There was also a significant effect of age group ($\chi^2(1) = 14.94$, $p < .001$), with older children more likely to give SO responses than younger children ($\beta = .88$; $z = 4.17$; $p < .001$).

**Table 8.11 Items with scorable response for comprehension and production: Experiment 2 (toys)**

| | | Comprehension | Production |
|---|---|---|---|
| 2½ yrs Dutch | SO | 94 (65.3%) | 115 (79.9%) |
| | OS | 50 (34.7%) | 29 (20.1%) |
| 3½ yrs Dutch | SO | 338 (88.0%) | 365 (95.1%) |
| | OS | 46 (12.0%) | 19 (4.9%) |

In Experiment 4, there were 89 items for 2½-year-olds and 234 for 3½-year-olds for which there was a scorable response for both comprehension and production. Table 8.12 shows how these items were answered by each age group. There was no significant interaction of task and age group $\chi^2(1) = 2.40$, $p > .1$). In the baseline model, there was a significant effect of task $\chi^2(1) = 48.20$, $p < .001$), with children more likely to give SO responses in production than comprehension ($\beta$ = .74; $z$ = 6.57; $p$ < .001). There was also a significant effect of age group ($\chi^2(1)$ = 4.95, $p$ = .03), with older children more likely to give SO responses than younger children ($\beta$ = .34; $z$ = 2.33; $p$ = .02).

**Table 8.12 Items with scorable response for comprehension and production: Experiment 4 (animations)**

| | | Comprehension | Production |
|---|---|---|---|
| 2½ yrs Dutch | SO | 55 (61.8%) | 72 (80.9%) |
| | OS | 34 (38.2%) | 17 (19.1%) |
| 3½ yrs Dutch | SO | 166 (70.9%) | 217 (92.7%) |
| | OS | 68 (29.1%) | 17 (7.3%) |

In Experiment 6, there were 190 items for 2½-year-olds and 136 for 3½-year-olds for which there was a scorable response for both comprehension and production. Table 8.13 shows how these items were answered by each age group. There was no significant interaction of task and age group $\chi^2(1)$ = .32, $p$ > .1). In the baseline model, there was a significant effect of task $\chi^2(1)$ = 20.86, $p$ < .001), with children more likely to give SO responses in production than comprehension ($\beta$ = .46; $z$ = 4.51; $p$ < .001). There was also a significant effect of age group ($\chi^2(1)$ = 7.93, $p$ <001), with older children more likely to give SO responses than younger children ($\beta$ = .53; $z$ = 3.00; $p$ = .003).

**Table 8.13 Items with scorable response for comprehension and production: Experiment 6 (animations)**

| | | Comprehension | Production |
|---|---|---|---|
| 2½ yrs English | SO | 120 (66.9%) | 150 (78.9%) |
| | OS | 70 (36.8%) | 40 (21.1%) |
| 3½ yrs English | SO | 109 (80.1%) | 125 (91.9%) |
| | OS | 27 (19.9%) | 11 (8.1%) |

In sum, all four experiments carried out with preschoolers had the same result: the older children were more likely to give SO responses than younger children, and all children were more likely to give SO responses in production than in comprehension. Despite all of the differences with regard to the influence of animacy across the different paradigms, age groups, languages, gaze patterns, etc., this one finding shines through consistently. Children produce word order in a more adult-like way than they comprehend it.

## 8.4 Discussion

The first analysis in this chapter looked at scorability. The comparison of scorability between the comprehension and production tasks showed that comprehension was more likely to be scorable than production. Adults showed variation in scorability in the production task only, and children were much more likely to answer scorably in comprehension than production. This supports the idea that elicited production tasks are intrinsically more noisy than the comprehension tasks used in these studies. Higher scorability on comprehension than production also suggests that an asymmetry of S-O word order cannot be explained by task effects, since elicited production seems not to be easier than act-out or picture selection tasks.

It seems safe to conclude that the production task was not easier than the comprehension task. There was not a more obvious but non-target form that many adults or children produced in the production task. While unscorable responses from one young child might be *Insufficient* ("cow . . . grass"), the unscorable responses from another child might be a confusion of the target verbs (when *push* is used to describe a *pull* event, this was conservatively scored as *Nontarget*). Other *Nontarget* responses included sentences like "The cow and the dog are walking" or "The duck is behind the airplane" but these were few and far between. In short, no clear alternative emerged in the child or adult unscorable utterances.

Did subject or object animacy play a role in scorability in either task? In comprehension, neither adults nor children were significantly affected by S-O animacy with regard to scorability. In production, on the other hand, there were some effects of animacy to be found. While adults of both languages produced scorable sentences over 90% of the time on average, they produced scorable sentences more often when the subject was animate. Subject and object animacy were not expected to influence how successfully a participant could produce a sentence with scorable word order since animations (and actions) were designed to be equally clear. In terms of accessibility, it could be that the retrieval of inanimate subjects is more demanding than the retrieval of animate subjects. Retrieval issues likely explain the fact that [+an +an] sentences were the least scorable for Dutch children in the Experiment 2b that used human toys, the names for which children sometimes had trouble retrieving.

The second analysis looked at accuracy. The comparison of accuracy on items that received a scorable response on both comprehension and production by an individual showed that adults generally responded with SO answers, regardless of task. Children, on the other hand, were more likely to respond with SO answers in production than in comprehension. This crucial result allows us to draw the conclusion that word order production exceeds word order comprehension in preschool speakers of Dutch and English.

With respect to age group, older children were—not surprisingly—more likely to respond with scorable actions, points, and utterances than their younger counterparts. In the elicitation task with toys in particular, the older children were better able to describe these demonstrated actions with a sentence with scorable word order. Older children were also more likely than younger children to respond with an SO response in either comprehension or production.

What this chapter demonstrates is that we can still draw conclusions about children's performance on comprehension and production in the face of inevitable data loss in production tasks. If the children in these studies see actions and describe them with sentences using SO word order, yet frequently ignore word order when interpreting the very same sentences, this strongly suggests that S-O word order functions more successfully in their grammars during sentence production than in comprehension. Whether or not S-O animacy is at the root of this asymmetry was looked at in the previous four chapters. All results taken together, in light of the theoretical and methodological issues previously addressed, are discussed in the next chapter.

# 9 Discussion

## An interpretation of the results

The preceding four chapters described six experiments, each with a comprehension and a production component. The set of experiments is unique in that young children's interpretation and use of word order was tested across three different task pairs. Experiments 1 and 2 tested Dutch preschoolers in an act-out task and an accompanying elicitation task. Experiments 3 – 6 tested Dutch- and English-speaking preschoolers and adult controls in a picture selection and an accompanying elicitation task. The preschoolers in experiments 4 and 6 were additionally tested with a preferential looking task. In the following discussion, the results of the entire set of experiments are interpreted as a whole.

This chapter is divided into three parts. I first revisit the two main research questions about asymmetry and animacy in light of the outcome of the experiments. Next, I re-address the theoretical account, noting where predictions were and were not met. Finally, I follow up on the matter of methodology, making some remarks about the tasks and materials that were used.

## 9.1 Outcome

In Chapter 1, we saw that there are several observed asymmetries in favor of production: children produce forms such as object pronouns, contrastive stress, and indefinite objects better than they comprehend them (Kuijper & Groothoff, 2010; Unsworth, 2007; de Villiers et al., 2006). Based on the results of the present study, can it be concluded that there is also an asymmetry between the production and comprehension of S-O word order? And if so, are there systematic mistakes in comprehension due to the relative animacy of subject and object across tasks? These two questions are answered in turn below.

### 9.1.1 Production-comprehension asymmetry

The results of Experiments 1 and 2 show that the Dutch preschoolers tested with act-out and elicitation tasks were able to produce S-O word order more reliably than they were able to comprehend it. Sentences were produced with SO order 95% – 100% of the time on average by the three-and-a-half-year-olds, whereas they had SO interpretations of the same sentences 84% – 91% of the time. The discrepancy appeared to be greater in the younger group of children: the two-and-a-half-year-olds produced SO order 81% – 85% of the time, but exhibited SO interpretations only about 59% – 64% of the time. The accuracy analysis presented in Chapter 8, which accounted for items that received a scorable response by a child on both comprehension and production, confirmed that task type (production vs. comprehension) was indeed a significant predictor of whether SO order would be reflected in a child's response. Namely, a child was more likely to produce word order as SO than he or she was to interpret word order as SO in the act-out paradigm.

Similar results were found in Experiments 4 and 6 with Dutch- and English-speaking preschoolers tested with picture selection and elicitation tasks. The average performance on the production task versus the picture selection task appeared to differ for both the older (92% vs. 70%) and younger (81% vs. 54%) Dutch children. Likewise, the average performance on the production task versus the picture selection task appeared to differ for the older (94% vs. 80%) and the younger (79% vs. 60%) English-speaking children. The accuracy analysis in Chapter 8 confirmed that task (production vs. comprehension) was again a significant predictor in both languages of whether SO order would be reflected in a child's response. A child was more likely to produce SO word order than he or she was to interpret word order as SO in the picture selection paradigm.

The results of the preferential looking task with animations in Experiments 4 and 6 reflected a general comprehension of word order by both age groups of preschoolers in both languages, since as the time progressed over the trial, their looks to the target animation increased. However, the proportion of looking to target never reached above 70% in any condition in either age group in either language. This maximum proportion of looks to SO animations was lower than the average performance on the accompanying production tasks with animations reported in the preceding paragraph.

It was reported in Chapter 8 that across the four experiments with preschoolers, a response in comprehension was more likely to be scorable than a response in production. However, scorability is separate from accuracy, so higher scorability in comprehension should not be interpreted as better performance in comprehension. On the contrary, the fact that at least two-thirds of the children in each experiment were able to produce mostly scorable responses in the face of an intrinsically open-ended elicitation task is seen as impressive. In other words, even though the elicitation tasks were more subject to noise than the act-out or picture selection tasks, children still produced SO sentences more often than they interpreted sentences as SO, evidenced by the accuracy analysis in Chapter 8. In sum, results of experiments 1, 2, 4, and 6 taken together are interpreted as evidence for an asymmetry in favor of production.

### 9.1.2 Subject-object animacy

The answer to the main question about whether there was an asymmetry between the comprehension and production of word order in this study is clear-cut. Determining the extent to which the relative animacy of subject and object affected the comprehension of word order across tasks is less straightforward. In what follows, I first summarize the effects of animacy exhibited by Dutch- and English-speaking adults on the picture selection task before turning to the less consistent effects of animacy exhibited by Dutch- and English-speaking preschoolers across three tasks. I will only discuss here the extent to which there were animacy effects; it is not until the next section that these will be held against the predictions of the theoretical model.

Dutch- and English-speaking adult controls performed similarly to each other on the picture selection task. An overview of the animacy effects in the experiments with adults is given in Table 9.1. Adults of both languages interpreted word order as SO 97% of the time on average, indicating a general preference to interpret initial NPs as subjects. However, they were more likely to do so when the subject was animate than when it was inanimate. Likewise, an effect of subject animacy was found in their reaction times, with participants of both languages faster to answer when the subject was animate. With regard to gaze patterns, adults of both languages showed increased looking to the target within three seconds of hearing the sentence subject. For Dutch adults only, this preference for the target was greater in sentences containing an animate subject compared to sentences with an inanimate subject. Thus, while word order was greatly adhered to, there were distinct effects of subject animacy in both the offline answers given by Dutch- and English-speaking adults, as well as in their processing behavior.

**Table 9.1 Overview of animacy effects in Experiments 3a and 5a**

| | Picture-selection task | | |
|---|---|---|---|
| | Accuracy | RT | Between-animation gaze |
| Dutch-speaking adults | preferred animate subjects | preferred animate subjects | preferred animate subjects |
| English-speaking adults | preferred animate subjects | preferred animate subjects | no effect |

Turning to preschoolers, the performance of the Dutch children was somewhat parallel across the act-out, picture selection, and preferential looking tasks. The three-and-a-half-year-olds consistently performed better than their younger two-and-a-half-year-old peers on the act-out and picture selection tasks. Regarding an effect of the animacy manipulation, the older Dutch children were also more likely to be affected by animacy of subject and/or animacy of object than their younger counterparts. The performance of the older children was affected by the animacy manipulation in the second act-out task (with a human-vehicle distinction) as well as in the picture-selection and looking tasks (both with an animal-vehicle distinction). The performance of younger children, on the other hand, was affected only in the picture selection task. No effect of animacy was found in either group on the act-out task with animal-vehicle distinction. An overview of the animacy effects in the experiments with Dutch preschoolers is given in the first part of Table 9.2.

The animacy effects exhibited by Dutch children are compatible with a preference for subjects to be animate and a preference for objects to be inanimate. The older Dutch children disprefered animate objects in the second act-out task and preferred animate subjects in the preferential looking task. The Dutch children of both age groups showed a combined preference for animate subjects and a dispreference for animate objects in the picture selection task. Taken together, these results can be interpreted as an effect of S-O animacy that spans across tasks, with the effects of animacy in the picture selection task appearing to be more complete than the effects of animacy in the act-out or preferential looking tasks.

**Table 9.2 Overview of animacy effects in Experiments 1a, 2a, 4a, & 6a**

| | Act-out (animals) | Act-out (human) | Picture selection (animals) | Preferential looking (animals) |
|---|---|---|---|---|
| Dutch 2½ | no effect | no effect | preferred animate subjects; dispreferred animate objects | no effect |
| Dutch 3½ | no effect | dispreferred animate objects | preferred animate subjects; dispreferred animate objects | preferred animate subjects |
| English 2½ | - | - | no effect | preferred animate objects |
| English 3½ | - | - | no effect | no effect |

Turning now to English-speaking preschoolers, we see quite a different picture. An overview of the animacy effects in the experiments with English preschoolers are given in the second part of Table 9.2. As was the case with the Dutch children, the older English-speaking children used word order more reliably than their younger counterparts in general on the picture selection task. The results differed in that both groups of Dutch children were affected by the animacy manipulation in this task, whereas neither group of English-speaking children was affected by the manipulation. Furthermore, of the two groups of Dutch speaking children, it was the older who showed a preference for animate subjects in the preferential looking study. Yet, of the two groups of English-speaking children, it was the younger who exhibited an effect of animacy. Namely, the English-speaking two-and-a-half-year-olds showed a greater increase in looks to the target when given sentences with animate objects. It can be concluded from this that the animacy effects on comprehension by the English-speaking children in this study were minimal compared to their Dutch peers.

To sum up the outcome of the entire set of experiments: there is evidence that preschool speakers of Dutch and English produce S-O word order better than they comprehend it. There is also evidence that the relative animacy of subject and of object influences the comprehension of all populations tested in this study. Adults of both languages showed a consistent preference to interpret animate entities as subject. Across tasks, Dutch preschoolers showed a general preference, most evident in the three-and-a-half-year-olds, to interpret animate entities as subject and inanimate entities as objects. The English-speaking children showed no such preference. The extent to which these results fit into the Optimality Theoretic framework introduced in Chapter 2 is discussed in the next section.

## 9.2 Theoretical issues

In Chapter 2, I summarized several ways variable comprehension or production could be accounted for. Explanations involving experimental artifacts or processing limitations were difficult to rule out since comprehension and production must be tested with different task and intrinsically involve different types of cognitive processing. When it comes to experimental artifacts, it might be argued that there were aspects of the present experimental design or materials that prevented a fair comparison between results on the comprehension and production tasks. However, the experiments in this study were controlled as much as possible (c.f. Chapter 3) in order to rule this type of explanation out. When it comes to processing limitations and the differences between the cognitive process of comprehending versus speaking, it might be argued that the children in this study had the challenge of figuring out what a speaker meant in comprehension, but had the luxury of assembling their own linguistic representations during production. However, this approach does not address the fact that children must also choose between different possible linguistic representations when speaking, nor does it account for examples in which the asymmetry goes in the opposite direction (e.g. problems with pronoun interpretation).

Optimality Theory (OT) was adopted as a framework for making predictions and for interpretting the results because it models comprehension and production as different directions of the same grammar. Hendriks, de Hoop, and Lamers' (2005) grammatical model of S-O word order interpretation for preschoolers was used as a starting point. Hendriks et al.'s model employs an animacy constraint that affects only interpretation (PROMINENCE) and a word order constraint that affects both interpretation and production (PRECEDENCE), and that the animacy constraint is incorrectly ranked above the word order constraint in the grammar of preschoolers. The model predicts that children learning Dutch and English will exhibit variable comprehension based on S-O animacy before coming to rely on word order. That is, children are expected to interpret S>O sentences as SO more often than S<O sentences due to their preference for S>O meanings, stemming from the animacy constraint.[19] On the basis Chan, Lieven, and

[19] The similarity between Optimality Theory and the Competition Model was addressed Chapter 2 (2.2.1.2). A key difference is that the Competition Model does not attempt to offer a formal model of the native speaker's underlying grammatical knowledge, or competence, in the

Tomasello's (2009) study of German and English, it was predicted that animacy effects would be found in the interpretations of English-speaking children at age 2;6, whereas these effects would persist in Dutch children of the age of 3;6. It was also predicted, following the real-time application of constraints proposed by Lamers and de Hoop (2004; 2005; de Hoop & Lamers, 2006) that effects of subject animacy could be found during the processing of the sentences in this study. Below, I show in what ways these predictions were met, followed by the ways they were not met and possible explanations for the discrepancies.

### 9.2.1 Predictions met

*Asymmetry between comprehension and production*

Children acquiring both English and Dutch were more likely to give SO responses in production than in comprehension, regardless of their age group. This was a consistent result, found in the four experiments carried out with preschoolers. The prediction that word order develops asymmetrically in children is met: children produce word order in a more adult-like way than they comprehend it.

*Variable comprehension in Dutch*

Of the three comprehension tasks carried out with Dutch preschoolers, the results of the picture selection task were most in line with predictions made. Both the two-and-a-half-year-old and the three-and-a-half-year-olds were influenced by S-O animacy in the picture selection task. In line with the predictions, Dutch children of both age groups interpreted word order well on S>O sentences (around 75% of the time) and they interpreted word order poorly when S<O (around 50% of the time). The performance of the three-and-a-half-year-olds on the second act-out task and the preferential looking task are also in line with predictions in that animate objects were dispreferred and animate subjects were preferred, respectively.

*Good comprehension by English-speaking preschoolers at 3;6*

It was expected that English-speaking preschoolers would rely on word order to interpret sentences by the age of three-and-a-half, which is what was found in the picture selection and preferential looking tasks. These children were able to select the target animation corresponding to a SO interpretation between 77% and 87% of the time across the sentence types in the picture selection task. The same group of children looked increasingly towards the target animations regardless of the type of sentence they heard in the preferential looking task.

*Effects of animacy in comprehension by adults*

Facilitation effects in adults' online processing of sentences with animate subjects were expected because sentences with animate subjects do not result in an early violation of PROMINENCE in an incremental model of grammatical processing. In both Dutch and English, facilitation effects were indeed found for these sentences in the form of faster average reaction times. But why would Dutch adults show an additional facilitation effect in looking behavior (i.e. greater increase of looking to the target when the subject was animate) not found in English-speaking adults? It was suggested in Chapter 6 that this is not incompatible with the OT model. Since it is proposed that Dutch places agreement, case, verb selection

---

way that OT does (Bates & MacWhinney, 1989: 32-36). The Competition Model is be able to explain the present comprehension results from a processing standpoint : constraints, or *cues*, are coming into conflict, resulting in at least a parsing failure rather than a true misinterpretation *per se*. However, OT simultaneously models the processing and the underlying knowledge, so a parsing failure can also be seen as a misinterpretation.

constraints above word order (De Hoop and Lamers 2006), while English does not, Dutch participants may be more sensitive to animacy information in the visual scene than English-speaking adults, even if an animacy constraint is ranked lower than a word order constraint in their grammar.

Although the offline effects of animacy additionally found in the languages (i.e. significantly greater accuracy on sentences with animate subjects) were not predicted, the result is not considered to be incompatible with the OT model. Participants of both languages overwhelmingly preferred SO interpretations (97%), seen as evidence that a word order constraint is ranked higher than an animacy constraint. At the same time, they were more likely to choose the target animation when the subject of the sentence was animate, seen as evidence that PROMINENCE is a constraint present in their grammar. Its influence may have become evident in offline responses since the experiment was very easy for adults. They tended to respond very quickly, perhaps at the cost of accuracy.

*SO order used in production*

In the elicitation tasks carried out with adult speakers of both Dutch and English, SO word order was produced 100% of the time. Furthermore, the Dutch preschoolers who participated in the act-out studies produced SO word order over 80% of the time, and at the same rate across the four sentence types, regardless of their age. This suggests that there is a preference to place subject before object and that differences in inherent animacy between subject and object do not trigger word order alternations in these populations. The fact that there were some effects of animacy found for adults during sentence production, as well as some word order alternations by Dutch- and English-speaking preschoolers in elicitation task that accompanied the picture selection tasks will be addressed in the next sub-section.

### 9.2.2 Predictions not met

*No variable comprehension in first act-out task*

The Dutch children of both age groups in Experiment 1a, tested with an act-out task, were unaffected by the animacy manipulation in sentences with an animal-vehicle distinction. This was the only experiment in which older Dutch preschoolers did not exhibit some form of animacy effect. It was first suggested that the distance between animals and vehicles on the animacy hierarchy might not be great enough to interfere with word order interpretation. However, the Dutch children of both age groups in Experiment 4a, tested with a picture selection task, were affected by the animacy manipulation in the same sentences with an animal-vehicle distinction. It cannot be that children are not sensitive to S-O animacy when tested with act-out tasks because the Dutch three-and-a-half-year-olds in the second act-out task (with human-vehicle distinction) as well as children in previous studies (Chapman & Miller, 1975; Chan, Lieven, & Tomasello, 2009) were influenced by animacy when tested with this task. If it is not the sentences themselves and not the task itself, it is not clear why the combination of animal-vehicle sentences and an act-out task resulted in fairly good interpretation of word order by the older children (around 90%) and fairly poor interpretation of word order by the younger children (around 60%) across the four sentence types. The anomalous findings of Experiment 1a demonstrate the importance of testing children on multiple tasks. The results of Experiment 1a taken alone would have led us to believe that Dutch three-and-a-half-year-olds are not affected by S-O animacy during sentence interpretation, when in fact the results of the remaining experiments suggest that they are.

*Inconsistent comprehension pattern of preschoolers at 2;6*

It was not only the first act-out task in which younger Dutch children seemed to be unaffected by the animacy manipulation. The results of all of the studies summarized in Table 9.2 show that both Dutch-and English-speaking two-and-a-half-year-olds were unaffected by the animacy manipulation, with two exceptions: the younger Dutch children showed an animacy effect in line with that of their older peers on the picture selection task, and the younger English-speaking children showed an unexpected preference for animate objects on the preferential looking task.

To begin with the Dutch children of two-and-a-half, the question is why they did not show an effect of animacy in either of the act-out tasks or the preferential looking task, since they did show one in the picture selection task. Regarding the act-out task, it could be that toy vehicles used in the task are salient to the children and this may have interfered with an animacy preference. However, Dutch children of this age also showed no preference during the looking task, carried out with the same materials that did trigger an influence of animacy when decision-making was involved (i.e. the picture selection task).

Regarding the English-speaking children of two-and-a-half, the question is why they did not show an effect of animacy in the picture selection task (whereas the Dutch children of the same age had), as well as why they would show an effect in the opposite direction than expected in the preferential looking task. This is a very strange result since young English-speaking preschoolers have been shown to be influenced by S-O animacy when interpreting word order in previous studies (Chapman & Miller, 1975; Chan, Lieven, & Tomasello, 2009). However, C&M did test children with a lower mean age than the children tested here (2;2 vs. 2;9), and they offered their items on different occasions to children until they got a scorable answer, which may have influenced their results. Chan et al. (2009) tested two-and-a-half-year-old children with novel verbs using a much greater distance in animacy between the entities (horse vs. telephone), which may have resulted in a clearer effect of animacy in their study than that found here. On the other hand, their “animacy” effect can also be seen as a plausibility effect if one considers the likelihood that an immobile entity such as a telephone can initiate an action, even if there is a special contraption involved enabling movement (described in Section 3.3.1.3). Thus, one could also conclude that English-speaking two-and-a-half-year-olds are simply not influenced by inherent animacy when tested with sentences that are truly reversible, whereas Dutch-speaking children of the same age are.

*Effects of animacy on production*

PROMINENCE is a constraint on comprehension that plays no role in production. Based on the distinction that can be made between contextual and inherent accessibility in lexical retrieval (Branigan, Pickering, & Tanaka, 2008; Tanaka, Branigan, & Pickering, 2005; van Bergen, 2011: 20-25), the prediction was made that the inherent accessibility of noun phrases is a very weak trigger of word order alternations. Only the animacy of the noun phrases in the sentences was manipulated, and not the discourse-level context, so no word order alternations were expected. That is, out of context, the nature of being a cow (as opposed to being a car) is not expected to result in the placement *cow* earlier in a sentence than *car*. This was largely the case, if we look at how often SO was used by Dutch and English-speaking adults in the control study, as well as the consistent use of SO across sentence types by the Dutch preschoolers in the elicitation tasks of the act-out paradigm.

On the other hand, both Dutch- and English-speaking preschoolers were affected by the animacy manipulation in the elicitation task where animations had to be described. Overall the children used SO order in their sentences between about 80% and 95% of the time, but for all age groups the Dutch children were more likely to do so when the agent was animate, and the English-speaking children were more likely to do so both when the agent was animate and when the patient was inanimate. This suggests that inherent accessibility might play a role in production for preschoolers, with a preference to place animate entities earlier in the sentence. That is, animate entities may by nature be easier for preschoolers to retrieve during production.

However, there is reason to believe this preference does not stem from the same animacy constraint that influences comprehension. In the case of Dutch children, performance on a comprehension task with animations showed an influence of both subject and object animacy that affected all sentence types, with boosted performance on S>O sentences and inhibited performance on S<O sentences. The same group’s performance on the elicitation task was extremely good in general, and the animacy effect was limited to subject animacy through a poorer use of SO order to describe animations with vehicle agents. A stronger case for the argument that animacy information is not treated the same way in comprehension as it is in production comes from the English-speaking children, who showed an effect of S-O animacy in production but not on the picture-selection task.

Turning now only to sentences that were produced with SO word order, there was a further real-time influence of S-O animacy on all populations’ general pattern of looking as they spoke. The general pattern is that they first look to agent, begin to speak, and then look to patient. Animacy appeared to affect this pattern in that, roughly speaking, children were inhibited by S<O animations and adults were facilitated by S<O animations. The older Dutch children showed a stronger switch of attention from agent to patient when describing animations with animate patients. The English-speaking preschoolers were also affected by S<O actions, with the greatest proportion of gaze to animate patients in animations with inanimate agent and animate patient both before and after beginning to speak. Dutch adults’ switch from agent to patient, on the other hand, was weakest in animations with animate agent and inanimate patient, and they were also faster to begin describing these animations. Similarly, English adults’ switch from agent to patient was weakest in animations with animate agents.

These online effects of animacy on gaze during production seem not to be the same kinds of effects of animacy on within-animation gaze when viewing animation pairs during comprehension. In the production task, though the agent-then-patient pattern may have been inhibited or facilitated by S-O prototypicality, participants of all ages still sought the agent first and then the patient regardless of their animacy. In the comprehension task (picture selection for adults and preferential looking for children), participants tended to look more at the most alive entity in their search for agent. This was evidenced by their general increased looking at agents in S=O animation pairs and their increased looking at animate entities in S≠O animation pairs. This was interpreted as a preference to find animate agents in comprehension and a preference simply to find agents in production.

While the search for agent in production may be facilitated when agents are animate or inhibited when agents are inanimate, a vehicle that is agent/subject in production is not disprefered in production to the same degree it is in comprehension (barring the asymmetry in the opposite direction found in English-speaking preschoolers). These results from the gaze analyses support the idea that animacy information has different effects in comprehension than in production.

In sum, the results generally support the OT account. First, the fact that there was an asymmetry between comprehension and production suggests that a model that incorporates these two directions of use of grammar is most appropriate. Second, the variable comprehension of word order found in Dutch preschoolers as well as animacy affects on Dutch- and English-speaking adult controls suggest that S-O animacy plays a systematic role in sentence interpretation due to an animacy constraint present in the grammar. However, variable comprehension was not found in all tasks carried out with preschoolers, with particularly unexpected results from English-speaking children and from children at age 2;6 in general. Furthermore, animacy played a role during the production task, indicating that animacy may be relevant in production, albeit in a different manner than in comprehension.

## 9.3 Methodological issues

Chapter 3 focused on methods of assessing production and comprehension of word order in young children. Sentence elicitation, act-out, picture selection, and preferential looking tasks were discussed in detail, as well as the controlled materials used in this study. In what follows I address some strengths and weaknesses of the experimental design, making some brief remarks about both the tasks and the materials.

### 9.3.1 Tasks

*Production*

In the first section of Chapter 3, the elicitation task was presented as the task most appropriate for testing word order production in a controlled setting. Comments by McClellan et al. (1986) led to the prediction that the elicited production task would be less engaging and more open-ended than the comprehension tasks. It was indeed more likely for a child to be excluded from a production analysis due to non-responses (lack of engagement) or unscorable responses (open-endedness) than from a comprehension analysis for the same reasons. However, about two-thirds of the children were enthusiastic about explaining the action or animation to the puppet and were able to complete the items, providing valuable data. Thus, the elicitation task is considered to have been a success in the face of inevitable noise.

Better use of word order in production than in comprehension is not seen as a result of having administered a comprehension task before the production task. Had the children been given the production task before the comprehension task, there would have been even more noise, i.e. unscorable responses. After all, including a control of first task for the adults showed that an adult participant who received production first was less likely to produce a scorable response than one who received comprehension first. Because the children were tested on comprehension and production in a brief 20-minute session, testing comprehension first was an excellent way to avoid unscorable responses by priming the children with simple, active sentences. Importantly, since the children heard the sentences during the comprehension task, but did not receive negative feedback if they indicated a reversed interpretation, word order was not taught.

A major contribution of this study is the finding that preschoolers can be successfully eye tracked during production. While preschoolers' voice onset latencies, or time it took to start describing the animation, contained too much noise to be analyzed in the way that adults' were, the gaze data just before and after they did start their description revealed an adult-like pattern. Dutch- and English-speaking adults and children tested in Experiments 3 – 6 adhered to the Griffin and Bock (2000) pattern: speakers first looked to the agent, and then they looked to the patient as they began their sentence. That is, the speaker sought out the entity that would be the active sentence's subject and then the entity that would be object. All English-speaking preschoolers and the older Dutch preschoolers showed a slight protraction in this pattern compared with adults, with a child's sentence beginning roughly 250 ms after they switched from looking at agent to looking at patient. The younger Dutch preschoolers, of which there were few left in the production analysis, showed a much greater protraction of about 750 ms. These results are in line with previous findings that children process sentences in the same way as adults, but may be slower (Sekerina, Stromswold, & Hestvik, 2004a; Yang et al., 2002)

*Comprehension*

In the second section of Chapter 3, several comprehension tasks appropriate for testing preschool-aged children were described: the act-out, picture selection, and preferential looking tasks. Both the act-out task and the picture selection task proved to be engaging since they required the children to actively and physically respond. The analysis in Chapter 8 showed that the scorability of responses on these comprehension tasks ranged between 93% and 100%. Therefore, these two tasks are seen as fun and effective comprehension tests for use with preschoolers.

Interestingly, in the act-out task in which Dutch children could perform any action upon hearing the test sentences, there was only an effect of object animacy found in the older children tested with human-vehicle sentences. But in general, the older Dutch children performed very well on both act-out tasks (about 80% – 90% SO interpretations) and the younger Dutch children had general difficulty (about 60% – 65% SO interpretations). It was only when children had the two alternate interpretations offered to them during the picture selection task that the clearest effect of S-O animacy was observed for Dutch (but not English-speaking) children of both age groups. These results suggest that Dutch children are most sensitive to animacy conflicts with word order in a task that literally pits the two interpretations against each other compared to a task that allows the children to demonstrate interpretations on their own.

In the third task, preferential looking, children were required to passively respond by simply watching the three-minute sequence of animation pairs. Key to the success of this task was its administration at the beginning of the session, after the pre-tests. It was at this point that the children were calmest and that the animations were freshest to them. Gaze contingent fixation crosses served to greatly diminish track loss, allowing all but a few children to be included in the preferential looking analyses. Thus, it is possible to effectively collect preferential gaze data from two-and-a-half- and three-and-a-half-year-olds using a remote eye tracker.

Although the looking task was predicted to be the easiest of the three tasks for young children since it requires the least from them, it is not apparent how it was able to reveal more about the word order comprehension of two-and-a-half-year-olds than the picture selection task. While these younger children showed an increase in looks to target animation over the four one-second windows of analysis, indicating that they generally looked to the animation that corresponded with a SO interpretation of the sentence they heard, the proportion of looks to the target animation task never reached above 60%. At the same time, this group of Dutch two-and-a-half-year-olds was able to choose SO interpretations over 70% of the time on average for the same S>O items in the picture selection task. In a similar fashion, the older Dutch children looked increasingly to the target animation after they heard the sentence, only reaching above 60% for S>O items, yet they were able to choose SO interpretations over 70% of the time for all sentence types save for S<O items in the picture selection task. In other words, if the looking task is able to reveal comprehension skills not always evident from a picture selection task, as suggested by Brandt-Kobele and Höhle (2010), it is not clear why children's use of word order—or the interfering effects of animacy—was not more evident in their preferential gaze data. The fact that an effect of S-O animacy was found during picture selection but not during preferential looking in a study carried out with precisely the same children and items suggests that preferential gaze may not be the most adequate measure for how children aged two-and-a-half or older are able to use word order and/or animacy information to interpret sentences.

### 9.3.2 Materials

In the third section of Chapter 3, experimental controls were discussed in detail. It was demonstrated how sentences with *push* and *pull* using either humans and vehicles or animals and vehicles are necessary to test only the effect of S-O animacy on word order. Because *push* and *pull* can take on animate or inanimate arguments and because no actions in the test sentences were stereotypical, the confounds of verb selectional restrictions and event likelihood were excluded. It seems that being restricted to *push* and *pull* was not problematic since any significant effects of verb were inconsistent. That is, a population may have performed better with *push* items or with *pull* items on a particular task, but it was usually the case that there was no significant effect of verb. An interesting observation is that children confused the two verbs in production (unfortunately resulting in an unscorable, non-target response) more frequently than they confused the two during the act-out task.

It was suggested in the previous section about comprehension tasks that the difference in the degree to which S-O animacy affected word order comprehension of Dutch children when tested with the act-out task versus picture selection may be attributed to the general difference in nature between the tasks: open interpretation vs. forced choice interpretation. However, there were also differences between the materials in these two tasks. Could it be that young children, boys in particular, find toy vehicles to be more salient than cartoon animations of vehicles? While there was no significant effect of gender in the responses of the children in the act-out tasks, an affinity for toy vehicles might explain the lack of consistent effects of animacy in the act-out tasks. After all, Dutch preschoolers tested in Experiment 4a with animations were sensitive to the same animacy contrasts as were tested in Experiment 1a with toys.

At the end of the day, the difficulty with testing the current research question in a completely controlled environment is that the inanimate entities must be capable of motion. This requires vehicles to be used with a target population, the male members of which tend to like playing with vehicles and whose culture currently embraces movies and cartoon series featuring anthropomorphized vehicles. However, I see this issue as less problematic than the inclusion of truly immobile entities as inanimate objects because the worse issue of plausibility is introduced. If a child chooses a frog as the agent when he hears the sentence *the pencil is pushing the frog*, how can we be sure whether it is a preference for animate agent or a preference for plausible agent? Thus in the current study, the fact that animacy effects were found with materials with animals vs. vehicles allows us to confidently conclude that preschoolers sometimes prefer animate agents and inanimate patients.

## 9.4 Summary of discussion

In this chapter, the outcome of the six experiments carried out in this study was discussed. Results were related to theoretical predictions and to methodological issues raised at the beginning of this dissertation. The Optimality Theoretic framework, which models production and comprehension as different directions of use of the same grammar, predicts asymmetry between comprehension and production of S-O word order by Dutch- and English-speaking preschoolers, as well as animacy effects on interpretation. These predictions were tested in a unique investigation involving three different comprehension tasks with accompanying production tasks.

In general, the results met the predictions of the OT models: we found an asymmetry between comprehension and production, variable comprehension of word order in Dutch preschoolers, and animacy effects on Dutch- and English-speaking adult controls. Systematic effects of animacy suggest that an animacy constraint is present in both child and adult grammar. Even though variable comprehension was not found in all tasks carried out with preschoolers (with particularly English-speaking children and younger Dutch-speaking children), the asymmetry in favor of production remained. And while animacy had an unexpected influence on word order production of English-speaking children, it did not influence their comprehension. This is seen as further evidence that the two directions of use of grammar do not operate identically.

We can be confident about the comprehension-production asymmetry found in this investigation because it occurred across different tasks that were adequately controlled for confounding factors, such as verb or likelihood biases. Children's comprehension of word order clearly suffers in the absence of other linguistic and contextual cues that normally aid them in determining agent and patient in natural situations. S-O animacy is rarely the only source of information about agent-patient relations, but may be used by preschoolers when everything else besides word order is stripped away. In contrast, children have little difficulty producing the same sentences that they have difficulty interpreting. The implications these results have for grammatical theory are addressed in the concluding chapter that follows.

# 10 Conclusion

## A real asymmetry

This dissertation presented an investigation of a particularly intriguing phenomenon: preschoolers fail to reliably use word order when interpreting reversible transitive sentences but they produce word order in an adult-like way. Even though it is more common for comprehension to precede and exceed production, the development of S-O word order appears to be an example of an asymmetry that goes in the opposite direction. Because previous studies that suggest that there is variable comprehension of word order by preschoolers did not always control for potentially confounding factors, such as the selectional restrictions of verbs or event likelihood, more evidence was needed before concluding that there is an asymmetry. The two central questions asked in the present research were whether there is an asymmetry between production and comprehension in early word order and whether comprehension is affected by the relative animacy of subject and object in a carefully controlled study.

Two theoretical models within the Optimality Theory framework were utilized to make specific predictions for early word order, as well as for adult word order, in Dutch and English. The first model, of Hendriks, de Hoop, and Lamers (2005), allowed for offline predictions. Their model predicts variable comprehension in preschoolers learning Dutch and English, whose grammar initially gives too much priority to an animacy constraint on meaning. Specifically, less SO interpretations are expected for sentences with an inanimate subject and animate object (S<O) than for other sentence types (S>O and S=O). Because the proposed animacy constraint does not help distinguish between forms to use in production, SO word order is predicted to be adhered to by preschoolers. The result is asymmetric development of early word order. Adult speakers of Dutch and English are proposed to have the word order constraint correctly ranked above the animacy constraint, so they are expected to interpret active, reversible sentences out of context as SO in comprehension and to use SO order in production.

The second model within the Optimality Theory framework applies the process of incremental optimization proposed by de Hoop and Lamers (2006). If constraints are evaluated as words are encountered in the linguistic stream, then sentences with inanimate subjects, especially S<O sentences, are predicted to be at a disadvantage during processing. In order to test the predictions of both models, six experiments were carried out, each with a comprehension and production component.

Based on the analysis in Chapter 8 and the discussion in Chapter 9, we can conclude from the four experiments carried out with child populations that there *is* an asymmetry between comprehension and production. In two experiments that used toys as stimuli with Dutch preschoolers, and two experiments that used cartoon animations as stimuli with Dutch and English preschoolers, children were more likely to produce word order as SO than they were to interpret word order as SO.

The results of the four experiments carried out with child populations together with the two experiments carried out with adult controls suggest that the relative animacy of subject and object influences the comprehension. When interpreting sentences that provided only word order and animacy information, adult speakers of Dutch and English consistently preferred to interpret animate entities as subject. Across tasks, Dutch preschoolers showed a general preference, most evident in the three-and-a-half-year-olds, to interpret animate entities as subject and inanimate entities as objects. However, the English-speaking children were not affected by animacy in comprehension, perhaps because they had acquired adult-like interpretation of word order already at age two-and-a-half.

The results of this investigation have shed light on the issue of asymmetry in child language. We now know that in a well-controlled study, preschoolers may not perform the same on a comprehension task as they will in production. For both Dutch- and English-speaking preschoolers, SO order was more likely to be used in production than in comprehension. Furthermore, this investigation has provided evidence that adult speakers of Dutch and English make use of animacy information in comprehension, both offline and during processing. These results have implications for acquisitional theory and for grammar in general.

The first implication of these results concerns how we define a grammar. If the different aspects of comprehension and production are delegated to the grammar, rather than delegated to general cognition, then the grammar itself can be seen as the root of asymmetry. Traditional approaches must appeal to extra-grammatical explanations for asymmetry, while Optimality Theory incorporates the different demands of comprehension and production into the grammar itself and challenges the traditional understanding of a grammar as a system of inviolable constraints. This eliminates the need to call upon extra-grammatical explanations for non-adult-like linguistic behavior of children. Furthermore, mis-ranked constraints as an explanation for the differences between child and adult language allows the same underlying mechanisms to be responsible for both systems—a decidedly more elegant approach than defining a whole new system (cf. Drenhaus & Fery, 2008). Thus, at an early point in language development, when it is said that children rank markedness constraints above faithfulness constraints, we are able to see evidence of an asymmetric grammar.

The second implication of these results of the present study concerns the role of animacy in the grammar. Is the effect of animacy linguistic or simply a general part of cognition? How much weight should animacy be given in universal grammar? The cross-linguistic evidence that animacy influences processing and has been grammaticalized in some languages is the motivation behind giving animacy an official role in grammar. The results of the present study, particularly the results from the adult speakers of Dutch and English, suggest that there is an underlying animacy constraint in universal grammar. It revealed itself not only through effects during linguistic processing, but also through effects in global sentence interpretation—this in languages that rank word order above animacy information. Thus, the incorporation of animacy into the grammar allows for a continuous, grammatical account of the acquisition of S-O word order.

# Appendix A

Experimental sentences

**Table A.1 Experimental sentences from Chapman and Miller (1975)**

| Type | Test sentence (A) | Type | Test sentence (B) |
|---|---|---|---|
| +an +an | The boy is hitting the girl<br>The girl is carrying the dog<br>The dog is chasing the boy | +an +an | The girl is hitting the boy<br>The dog is carrying the girl<br>The boy is chasing the dog |
| +an -an | The dog is chasing the car<br>The boy is carrying the truck<br>The girl is pulling the boat | -an +an | The car is chasing the dog<br>The truck is carrying the boy<br>The boat is pulling the girl |
| -an +an | The boat is hitting the girl<br>The truck is bumping the dog<br>The car is pushing the boy | +an -an | The girl is hitting the boat<br>The dog is bumping the truck<br>The boy is pushing the car |
| -an -an | The truck is pulling the boat<br>The boat is bumping the car<br>The car is pushing the truck | -an -an | The boat is pulling the truck<br>The car is bumping the boat<br>The truck is pushing the car |

**Table A.2 Overview of experimental sentences**

| | Type | Test sentence (A) | Type | Test sentence (B) |
|---|---|---|---|---|
| 1 | [+ +] | De koe duwt de hond[1] | [+ +] | De hond duwt de koe |
| | | *The cow is pushing the dog* | | *The dog is pushing the cow* |
| 2 | [ - +] | De auto duwt de koe | [+ - ] | De koe duwt de auto |
| | | *The car is pushing the cow* | | *The cow is pushing the car* |
| 3 | [+ - ] | De hond duwt de trein/bus[2] | [ - +] | De trein/bus duwt de hond |
| | | *The dog is pushing the train/bus* | | *The train/bus is pushing the dog* |
| 4 | [ - - ] | De trein/bus duwt de auto | [ - - ] | De auto duwt de trein/bus |
| | | *The train/bus is pushing the car* | | *The car is pushing the train/bus* |
| 5 | [+ +] | De eend duwt de vis | [+ +] | De vis duwt de eend |
| | | *The duck is pushing the fish* | | *The fish is pushing the duck* |
| 6 | [ - +] | Het vliegtuig duwt de eend | [+ - ] | De eend duwt het vliegtuig |
| | | *The airplane is pushing the duck* | | *The duck is pushing the airplane* |
| 7 | [+ - ] | De vis duwt de boot | [ - +] | De boot duwt de vis |
| | | *The fish is pushing the boat* | | *The boat is pushing the fish* |
| 8 | [ - - ] | De boot duwt het vliegtuig | [ - - ] | Het vliegtuig duwt de boot |
| | | *The boat is pushing the airplane* | | *The airplane is pushing the boat* |
| 9 | [+ +] | De hond trekt de koe | [+ +] | De koe trekt de hond |
| | | *The dog is pulling the cow* | | *The cow is pulling the dog* |
| 10 | [+ - ] | De koe trekt de auto | [ - +] | De auto trekt de koe |
| | | *The cow is pulling the car* | | *The car is pulling the cow* |
| 11 | [ - +] | De trein/bus trekt de hond | [+ - ] | De hond trekt de trein |
| | | *The train/bus is pulling the dog* | | *The dog is pulling the train/bus* |
| 12 | [ - - ] | De auto trekt de trein/bus | [ - - ] | De trein/bus trekt de auto |
| | | *The car is pulling the train/bus* | | *The train/bus is pulling the car* |
| 13 | [+ +] | De vis trekt de eend | [+ +] | De eend trekt de vis |
| | | *The fish is pulling the duck* | | *The duck is pulling the fish* |
| 14 | [+ - ] | De eend trekt het vliegtuig | [ - +] | Het vliegtuig trekt de eend |
| | | *The duck is pulling the airplane* | | *The airplane is pulling the duck* |
| 15 | [ - +] | De boot trekt de vis | [+ - ] | De vis trekt de boot |
| | | *The boat is pulling the fish* | | *The fish is pulling the boat* |
| 16 | [ - - ] | Het vliegtuig trekt de boot | [ - - ] | De boot trekt het vliegtuig |
| | | *The airplane is pulling the boat* | | *The boat is pulling the airplane* |

1. In Experiment 1b: *koe*, *hond*, *vis*, and *eend* (*cow*, *dog*, *fish*, *duck*) were replaced by *mevrouw*, *meisje*, *jongen*, and *meneer* (*lady*, *girl*, *boy*, *gentleman*) respectively.
2. Experiments with toys (1 & 2) used *train*, Experiments with pictures/videos (4-6) used *bus*.

**Table A.3 Vocabulary norms for words in experimental sentences (30 months)**

| [[Type of item | Dutch | Percent who understand and say word[1] | English | Percent who understand and say word[2] |
|---|---|---|---|---|
| Test | koe | 100.00 | cow | 94.30 |
| | hond(je) | 100.00 | dog | 97.10 |
| | vis | 96.72 | fish | 95.70 |
| | eend | 95.08 | duck | 95.70 |
| | auto | 100.00 | car | 98.60 |
| | bus | 96.72 | bus | 94.30 |
| | boot | 98.36 | boat | 95.70 |
| | vliegtuig | 95.08 | airplane | 97.10 |
| | meneer | 83.61 | (man)[3] | 88.60 |
| | mevrouw | 80.33 | (lady) | 78.60 |
| | jongen | 81.97 | (boy) | 94.30 |
| | meisje | 81.97 | (girl) | 92.90 |
| | duwen | 78.69 | push | 81.40 |
| | trekken | 75.41 | pull | 74.30 |
| Practice/filler | beer | 96.72 | bear | 95.70 |
| | kikker | 85.25 | frog | 90.00 |
| | poes(je) | 98.36 | cat | 92.90 |
| | aap | 88.52 | monkey | 91.40 |
| | konijn(tje) | 96.72 | bunny/rabbit | 92.90 |
| | schaap | 88.52 | (sheep) | 82.90 |
| | paard | 96.72 | (horse) | 98.60 |
| | varken | 91.80 | (pig) | 92.90 |
| | appel | 100.00 | apple | 98.60 |
| | bal | 100.00 | ball | 100.00 |
| | water | 98.36 | water | 100.00 |
| | ballon | 98.36 | balloon | 100.00 |
| | fles | 95.08 | bottle | 95.70 |
| | stoel | 96.72 | chair | 97.10 |
| | fiets | 100.00 | bicycle/bike | 98.60 |
| | blok | 95.08 | (block) | 94.30 |
| | brommer | 91.80 | (motorcycle) | 80.00 |
| | helikopter | 80.33 | (helicopter) | 84.30 |
| | springen | 90.16 | jump | 92.90 |
| | vallen | 91.80 | fall | 98.60 |
| | dragen | 78.69 | carry | 84.30 |
| | kietelen | 72.13 | tickle | 90.00 |
| | rijden | 85.25 | ride | 90.00 |
| | schoppen | 80.33 | kick | 90.00 |
| | wijzen | 80.33 | point | na |
| | rollen | na | roll | na |
| | stuiteren | na | bounce | na |
| | kussen | 90.16 | (kiss) | 98.60 |
| | vliegen | na | (fly) | na |
| | liggen | na | (lie) | na |

1. Normed data for Dutch comes from N-CDI (Zink & Lejaegere, 2002).
2. Normed data for English comes from CDI (Dale & Fenson, 1996).
3. English words occuring in parentheses were not used in the experiment with English-speaking children.

**Table A.4 Practice and filler sentences from Experiments 1 & 2**

| | Item | Type | Sentence[1] |
|---|---|---|---|
| Comprehension | Practice | [+ +] | de kikker/koe springt over het varken[2]<br>*the frog/cow is jumping over the pig* |
| | Practice | [+ - ] | het varken kust het blokje<br>*the pig is kissing the block* |
| | Practice | [ - +] | de brommer valt op het paard<br>*the moped is falling on the horse* |
| | Practice | [ - +] | de helikopter vliegt over het schaap<br>*the helicopter is flying over the sheep* |
| | Practice | [ - - ] | de helikopter vliegt over het blokje<br>*the helicopter is flying over the block* |
| | Filler | [+ +] | het schaap kust het varken<br>*the sheep is kissing the pig* |
| | Filler | [+ - ] | de kikker/koe springt over de brommer<br>*the frog/cow is jumping over the moped* |
| | Filler | [ - +] | het blokje valt op het schaap<br>*the block is falling on the sheep* |
| | Filler | [ - - ] | de brommer rijdt over het blokje<br>*the moped is riding over the block* |
| Production | Practice | [+ +] | het schaap kust het varken*<br>*the sheep is kissing the pig* |
| | Practice | [+ +] | het varken springt over de kikker/koe<br>*the pig is jumping over the frog/cow* |
| | Practice | [+ - ] | het paard springt over de brommer<br>*the horse is jumping over the moped* |
| | Practice | [ - +] | het blokje valt op het schaap*<br>*the block is falling on the sheep* |
| | Practice | [ - - ] | de helikopter vliegt over het blokje*<br>*the helicopter is flying over the block* |

1. A sentence marked with * also occurs as a comprehension practice or filler item.
2. *Kikker* (*frog*) in Experiment 1 is replaced with *koe* (*cow*) in Experiment 2.

**Table A.5 Practice and filler sentences from Experiments 3 & 5 (Dutch & English adults)**

| | Item | Type | Target sentence[1] |
|---|---|---|---|
| Comprehension | Practice | [+ +] | de aap kietelt de beer<br>*the monkey is tickling the bear* |
| | Practice | [ - +] | de bal vliegt over de poes<br>*the ball is flying/bouncing over the cat* |
| | Practice | [ - - ] | de ballon draagt de fles<br>*the ballon is carrying the bottle* |
| Production | Practice | [+ +] | de aap kietelt de beer*<br>*the monkey is tickling the bear* |
| | Practice | [+ - ] | de konijn bijt de bal<br>*the rabbit is biting the ball* |
| | Practice | [ - - ] | de ballon draagt de fles*<br>*the balloon is carrying the bottle* |
| Comprehension & Production | Filler | [+ +] | de geit bijt het varken<br>*the goat is biting the pig* |
| | Filler | [+ +] | de haai bijt de gans<br>*the shark is biting the goose* |
| | Filler | [+ +] | de kikker springt over de schildpad<br>*the frog is jumping over the turtle* |
| | Filler | [+ +] | het konijn springt over de poes<br>*the rabbit is jumping over the cat* |
| | Filler | [+ - ] | de gans vliegt over het vrachtschip<br>*the goose is flying over the freightship* |
| | Filler | [+ - ] | de poes springt over de bakfiets<br>*the cat is jumping over the bike* |
| | Filler | [+ - ] | de schildpad bijt de plank<br>*the turtle is biting the board* |
| | Filler | [+ - ] | het varken bijt de bal<br>*the pig is biting the ball* |
| | Filler | [ - +] | de plank wiebelt op de haai<br>*the board is wobbling on the shark* |
| | Filler | [ - +] | de vrachtauto vervoert de kikkers<br>*the truck is transporting the frogs* |
| | Filler | [ - +] | het poortje slaat tegen de geit<br>*the gate is hitting the goat* |
| | Filler | [ - +] | het vrachtschip vervoert de konijnen<br>*the freightship is transporting rabbits* |
| | Filler | [ - - ] | de bakfiets vervoert de flessen<br>*the bike is transporting bottles* |
| | Filler | [ - - ] | de bal rolt langs het poortje<br>*the ball is rolling past the gate* |
| | Filler | [ - - ] | de fles wiebelt op de plank<br>*the bottle is wobbling on the board* |
| | Filler | [ - - ] | de trein vervoert de planken<br>*the train is transporting the board* |

1. A sentence marked with * also occurs as a comprehension practice or filler item.

**Table A.6 Practice and filler sentences from Experiments 4 & 6 (Dutch & English preschoolers)**

| | Item | Type | Target sentence[1] |
|---|---|---|---|
| Comprehension | Practice | [+ +] | de aap kietelt de beer<br>*the monkey is tickling the bear* |
| | Practice | [+ +] | het konijn springt over de poes<br>*the rabbit is jumping over the cat* |
| | Practice | [ - +] | de bal rolt op de poes<br>*the ball is rolling past the cat* |
| | Practice | [ - +] | de bal vliegt over de poes<br>*the ball is flying over the cat* |
| | Practice | [ - - ] | de ballon draagt de fles<br>*the balloon is carrying the bottle* |
| | Practice | [ - - ] | de fles ligt op het water<br>*the bottle is (lying) on the water* |
| | Filler | [+ +] | de poes springt over het konijn<br>*the cat is jumping over the rabbit* |
| | Filler | [+ - ] | de beer schopt tegen de fiets<br>*the bear is kicking the bike* |
| | Filler | [ - +] | de appel valt op de kikker<br>*the apple is falling on the frog* |
| | Filler | [ - - ] | de bal rolt onder de stoel<br>*the ball is rolling under the chair* |
| Production | Practice | [+ +] | de kikker springt over de poes<br>*the frog is jumping over the cat* |
| | Practice | [+ +] | de poes springt over het konijn<br>*the cat is jumping over the rabbit* |
| | Practice | [+ - ] | de beer rijdt op de fiets<br>*the bear is riding the bike* |
| | Practice | [+ - ] | de konijn bijt de bal<br>*the rabbit is biting the ball* |
| | Practice | [ - +] | de bal rolt op de poes*<br>*the ball is rolling on the cat* |
| | Practice | [ - - ] | de ballon draagt de bal<br>*the balloon is carrying the ball* |

1. A sentence marked with * also occurs as a comprehension practice or filler item.

# Appendix B

Experimental toys and animations

## Experiments 1 & 2: Toys

**Figure B.1 Toys used in test items of Experiment 1**

**Figure B.2 Toys used in practice and filler items of Experiment 1**

**Figure B.3 Toys used in verb demonstration of Experiment 1**

**Figure B.4 Toys used in test items of Experiment 2**

**Figure B.5 Toys used in filler and practice items Experiment 2**

**Figure B.6 Toys used in verb demonstration of Experiment 2**

# Experiments 3 – 6: Animations

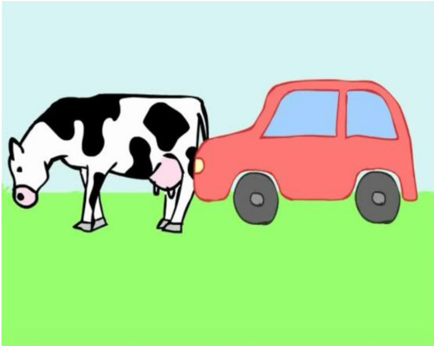

**Figure B.7 The car is pushing the cow**

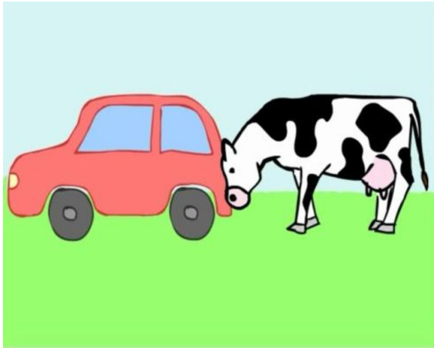

**Figure B.8 The cow is pushing the cow**

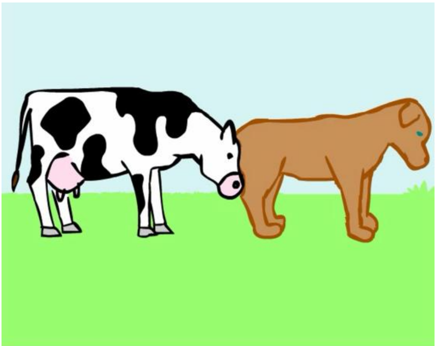

**Figure B.9 The cow is pushing the dog**

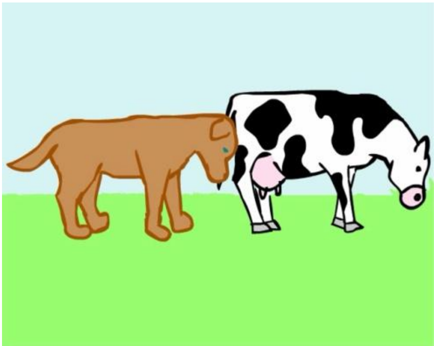

**Figure B.10 The dog is pushing the cow**

**Figure B.11 The dog is pushing the bus**

**Figure B.12 The bus is pushing the dog**

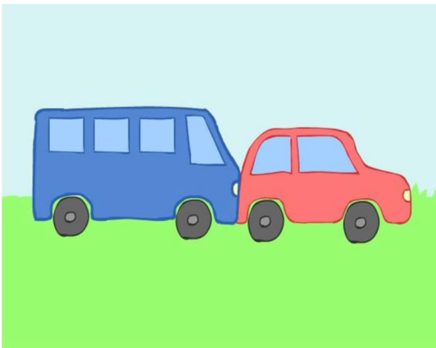

**Figure B.13 The bus is pushing the car**

**Figure B.14 The car is pushing the bus**

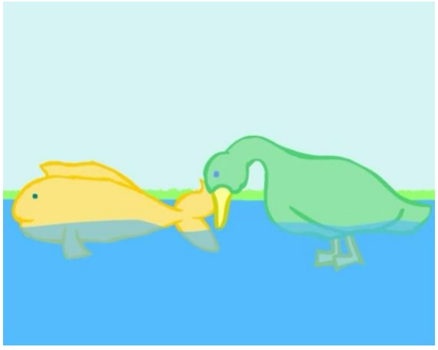

**Figure B.15 The duck is pushing the fish**

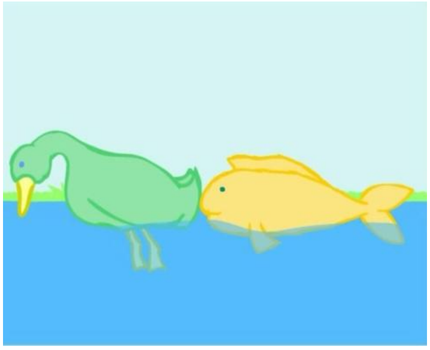

**Figure B.16 The fish is pushing the duck**

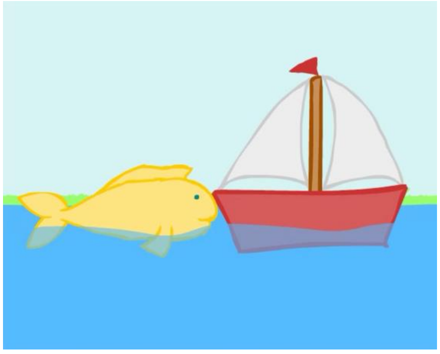

**Figure B.17 The fish is pushing the boat**

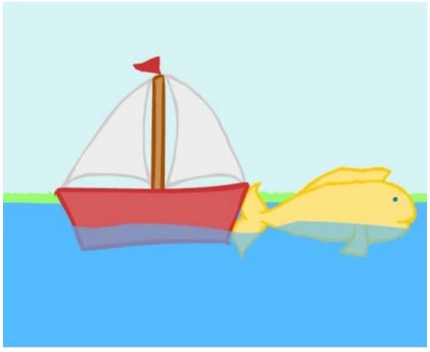

**Figure B.18 The boat is pushing the duck**

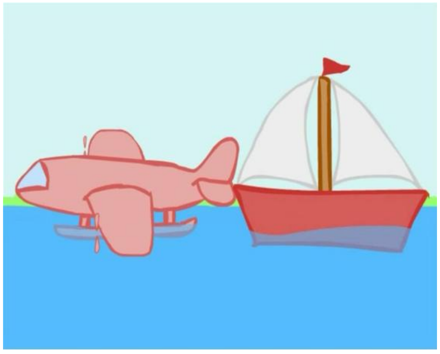

**Figure B.19 The boat is pushing the airplane**

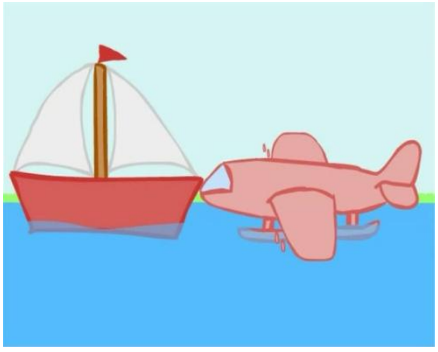

**Figure B.20 The airplane is pushing the boat**

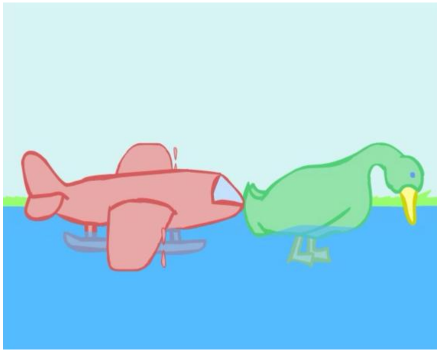

**Figure B.21 The airplane is pushing the duck**

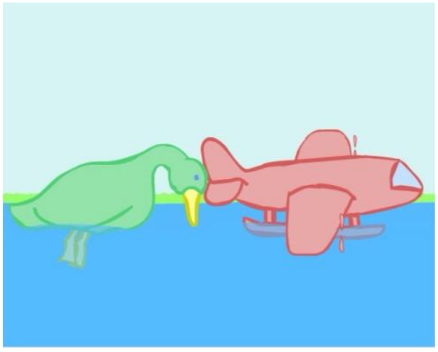

**Figure B.22 The duck is pushing the airplane**

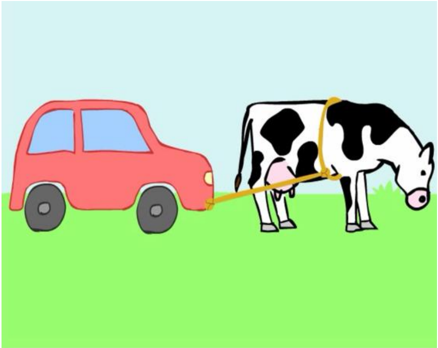

**Figure B.23 The cow is pulling the car**

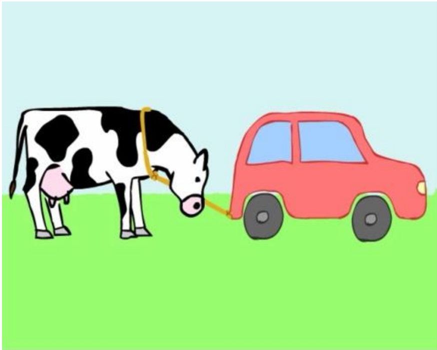

**Figure B.24 The car is pulling the cow**

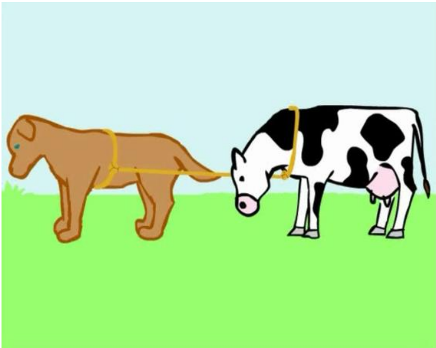

**Figure B.25 The dog is pulling the cow**

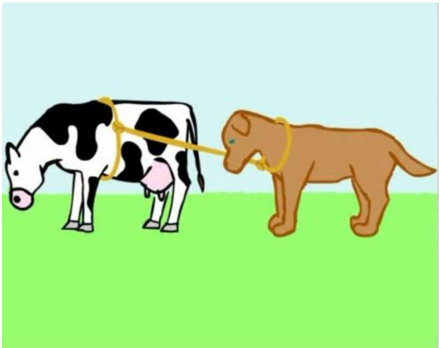

**Figure B.26 The cow is pulling the dog**

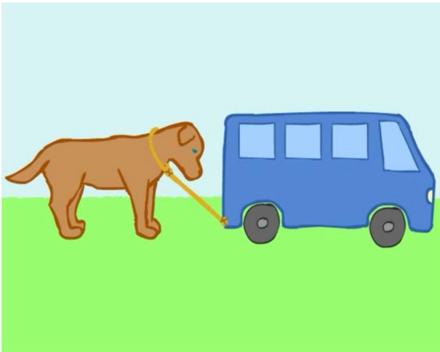

**Figure B.27 The bus is pulling the dog**

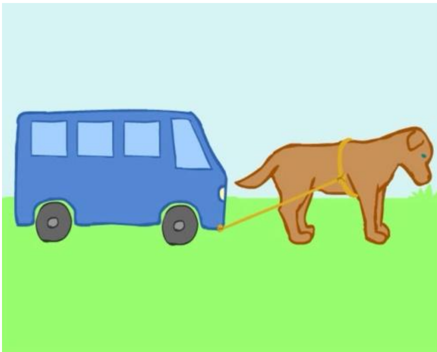

**Figure B.28 The dog is pulling the bus**

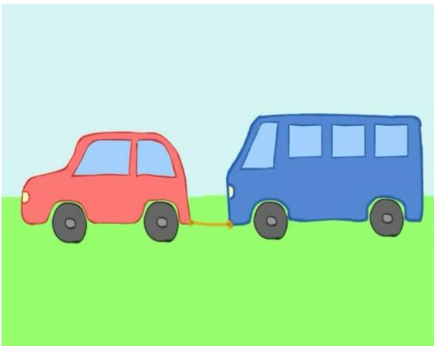

**Figure B.29 The car is pulling the bus**

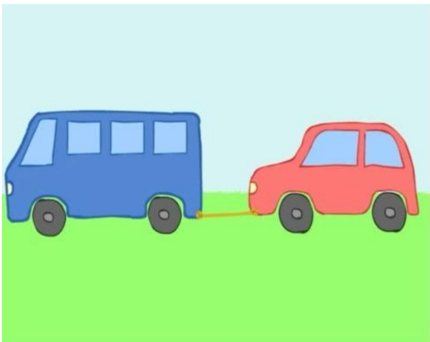

**Figure B.30 The bus is pulling the car**

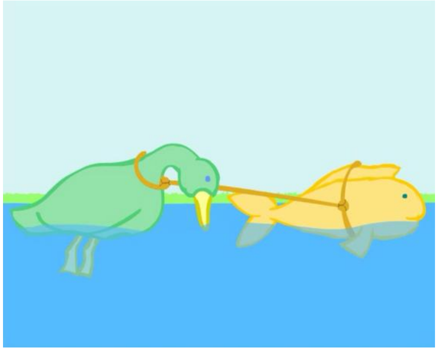

**Figure B.31 The fish is pulling the duck**

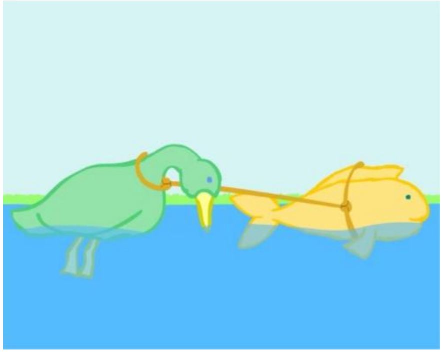

**Figure B.32 The fish pulls the duck**

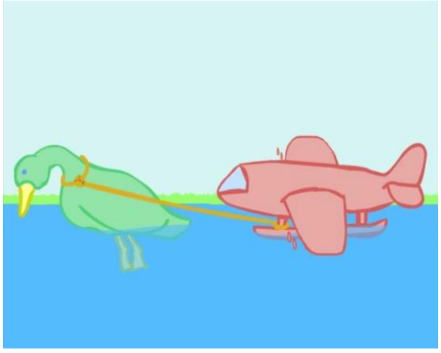

**Figure B.33 The duck is pulling the airplane**

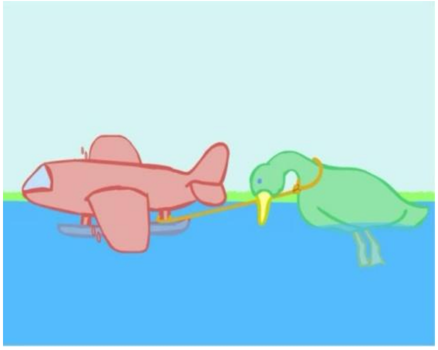

**Figure B.34 The airplane is pulling the duck**

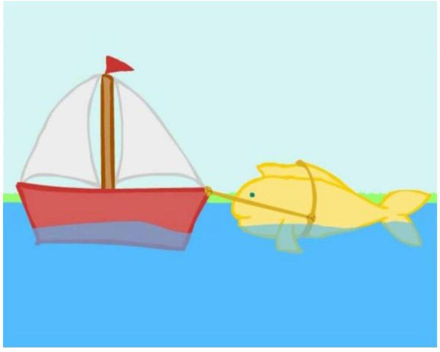

**Figure B.35 The boat is pulling the fish**

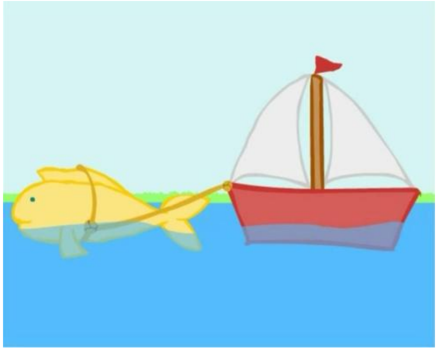

**Figure B.36 The fish is pulling the boat**

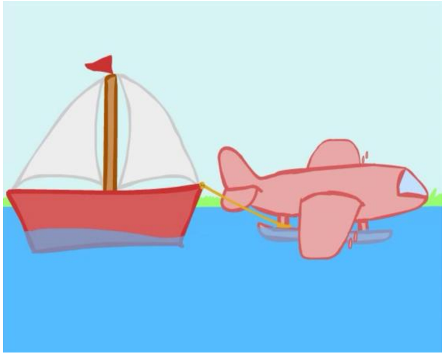

**Figure B.37 The airplane is pulling the boat**

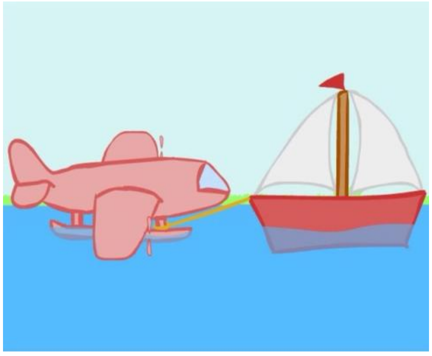

**Figure B.38 The boat is pulling the airplane**

# Appendix C

Preschoolers' individual scores

# Experiments 1a & 2a (Chapter 4)

**Table C.1 Individual comprehension scores per sentence type: Experiment 1a (2½ yrs)**

| | Gender | Age | N-CDI percentile | Number of SO interpretations* | | | |
|---|---|---|---|---|---|---|---|
| | | | | [++] | [+ -] | [- +] | [- -] |
| 1 | F | 2;5 | .10 | 2 | 2 | 1 | 1 |
| 2 | M | 2;7 | .15 | 3 | 4 | 3 | 1 |
| 3 | M | 2;7 | .85 | 0/3 | 2 | 4 | 2 |
| 4 | M | 2;7 | .95 | 0/2 | 2/3 | 3/3 | 1/3 |
| 5 | M | 2;7 | .46 | 1/2 | 1/2 | 2/2 | 2/3 |
| 6 | M | 2;11 | .65 | 2 | 3 | 1 | 3 |
| 7 | M | 2;11 | .50 | 3 | 2 | 4 | 3 |
| 8 | F | 3;0 | .70 | 3 | 2 | 4 | 4 |
| 9 | F | 3;1 | .10 | 1 | 4 | 2 | 2 |
| 10 | F | 3;1 | .70 | 2 | 2/3 | 2/3 | 3 |
| 11 | F | 3;2 | .63 | 4 | 3 | 3 | 0 |
| 12 | M | 3;2 | .45 | 2 | 2 | 2 | 2 |

*Out of 4 unless otherwise noted

**Table C.2 Individual comprehension scores per sentence type: Experiment 1a (3½ yrs)**

| | Gender | Age | KINT score (%) | Number of SO interpretations* | | | |
|---|---|---|---|---|---|---|---|
| | | | | [++] | [+ -] | [- +] | [- -] |
| 13 | M | 3;4 | 22 | 4 | 4 | 4 | 4 |
| 14 | F | 3;4 | 42 | 3 | 3 | 4 | 3/3 |
| 15 | M | 3;5 | 27 | 4 | 4 | 4 | 4 |
| 16 | F | 3;5 | 85 | 2 | 2 | 3 | 3 |
| 17 | F | 3;6 | 45 | 4 | 4 | 4 | 4 |
| 18 | M | 3;6 | 54 | 2 | 4 | 4 | 4 |
| 19 | F | 3;8 | 39 | 4 | 4 | 4 | 4 |
| 20 | M | 3;9 | 98 | 4 | 4 | 4 | 4 |
| 21 | M | 3;9 | 50 | 4 | 3 | 4 | 4 |
| 22 | M | 3;9 | 45 | 3 | 4 | 4 | 4 |
| 23 | F | 3;9 | 59 | 4 | 4 | 4 | 4 |
| 24 | F | 3;10 | 47 | 4 | 4 | 4 | 3 |
| 25 | M | 3;10 | 42 | 3 | 4 | 2 | 3 |
| 26 | F | 3;10 | 61 | 4 | 3/3 | 2/3 | 3/3 |
| 27 | F | 3;10 | 61 | 4 | 4 | 4 | 4 |
| 28 | M | 3;11 | 67 | 4 | 4 | 4 | 4 |

*Out of 4 unless otherwise noted

**Table C.3 Individual comprehension scores per sentence type: Experiment 2a (2½ yrs)**

| Gender | Age | Number of SO interpretations* | | | |
|---|---|---|---|---|---|
| | | [++] | [+ -] | [- +] | [- -] |
| M | 2;6 | 1/3 | 0 | 4 | 1 |
| M | 2;7 | 1 | 1 | 3 | 3 |
| F | 2;8 | 2 | 3 | 1/3 | 2 |
| F | 2;8 | 1 | 2 | 3 | 2 |
| F | 2;8 | 2 | 2 | 3 | 1 |
| M | 2;8 | 1/2 | 1/2 | 2/2 | 2/2 |
| M | 2;8 | 4 | 3 | 3/3 | 2 |
| M | 2;8 | 1 | 4/3 | 1/3 | 2 |
| F | 2;9 | 4 | 3/3 | 2 | 2 |
| F | 2;9 | 1/2 | 1/3 | 2/3 | 1/3 |
| F | 2;9 | 4 | 3 | 3 | 4 |
| F | 2;9 | 1 | 1 | 4 | 1 |
| M | 2;9 | 2 | 2 | 4 | 4 |
| M | 2;9 | 1 | 3 | 3 | 1 |
| M | 2;10 | 4 | 3 | 4 | 4 |
| M | 2;10 | 2/2 | 2/2 | 0/2 | 2/3 |
| F | 2;11 | 4 | 4 | 2 | 3 |
| M | 2;11 | 2 | 2 | 4 | 2 |
| M | 2;11 | 3 | 1 | 4 | 3 |
| M | 2;11 | 4 | 1 | 3 | 3 |
| F | 3;0 | 4 | 4 | 1 | 4 |
| M | 3;0 | 3 | 3 | 4 | 2 |
| F | 3;1 | 1/2 | 2/2 | 0/2 | 2/2 |
| F | 3;2 | 2 | 3 | 2 | 1 |
| F | 3;2 | 3 | 4 | 2 | 2 |
| F | 3;2 | 3 | 0 | 4 | 3 |
| F | 3;2 | 2 | 2 | 2 | 2 |
| M | 3;2 | 2 | 2 | 2 | 3/3 |
| M | 3;2 | 2 | 4 | 1 | 3 |
| M | 3;2 | 4 | 2 | 4 | 4/3 |

*Out of 4 unless otherwise noted

**Table C.4 Individual comprehension scores per sentence type for: Experiment 2a (3½ yrs)**

| Gender | Age | Number of SO interpretations* | | | |
|---|---|---|---|---|---|
| | | [++] | [+ -] | [- +] | [- -] |
| F | 3;3 | 4 | 4 | 4 | 4 |
| F | 3;3 | 4 | 4 | 4 | 4 |
| F | 3;3 | 4 | 4 | 4 | 4 |
| M | 3;3 | 3 | 4 | 3 | 3/3 |
| M | 3;3 | 3 | 1 | 4 | 2 |
| M | 3;3 | 3 | 4 | 0 | 2 |
| M | 3;3 | 4 | 4 | 4 | 4 |
| F | 3;4 | 4 | 3 | 4 | 4 |
| F | 3;4 | 2/3 | 2/3 | 3 | 4 |
| F | 3;4 | 2/3 | 3 | 0/2 | 3/3 |
| M | 3;4 | 3 | 3 | 4 | 4 |
| F | 3;5 | 4 | 4 | 0 | 3 |
| M | 3;5 | 4 | 4 | 4 | 4 |
| M | 3;5 | 3 | 4 | 4 | 4 |
| M | 3;5 | 3 | 3 | 3 | 3 |
| M | 3;5 | 4 | 4 | 4 | 4 |
| F | 3;6 | 2 | 1 | 3/3 | 2 |
| M | 3;6 | 3 | 4 | 4 | 4 |
| M | 3;6 | 3 | 2 | 3 | 3 |
| F | 3;7 | 3 | 4 | 3 | 4 |
| F | 3;7 | 4 | 4 | 1 | 4 |
| F | 3;7 | 3 | 3 | 1 | 2 |
| M | 3;7 | 4 | 4 | 4 | 4 |
| F | 3;8 | 4 | 4 | 4 | 4 |
| M | 3;8 | 4 | 4 | 4 | 4 |
| M | 3;8 | 0/2 | 2/2 | 1/2 | 1/2 |
| M | 3;8 | 4 | 3 | 2/3 | 3 |
| M | 3;8 | 3 | 4 | 4 | 3 |
| F | 3;9 | 2 | 4 | 4 | 4 |
| F | 3;9 | 3 | 4 | 4 | 4 |
| M | 3;9 | 3 | 3 | 2 | 2 |
| F | 3;10 | 3 | 4 | 4 | 4 |
| M | 3;10 | 4 | 4 | 3 | 4 |

*Out of 4 unless otherwise noted

# Experiments 1b & 2b (Chapter 5)

**Table C.5 Individual production scores per sentence type: Experiment 1b (2½ yrs)**

| Gender | Age | N-CDI percentile | Number of SO order productions* | | | |
|---|---|---|---|---|---|---|
| | | | [++] | [+ -] | [- +] | [- -] |
| M | 2;7 | .95 | 2 | 3/3 | 3 | 1/2 |
| M | 2;7 | .85 | 4 | 3 | 3/3 | 2 |
| M | 2;11 | .50 | 3 | 3 | 4 | 4 |
| F | 3;0 | .70 | 2/2 | 2/2 | 2/2 | 1/2 |
| M | 3;2 | .45 | 3/3 | 3/3 | 2/2 | 3/3 |

*Out of 4 unless otherwise noted

**Table C.6 Individual production scores per sentence type: Experiment 1b (3½ yrs)**

| Gender | Age | KINT score (%) | Number of SO order productions* | | | |
|---|---|---|---|---|---|---|
| | | | [++] | [+ -] | [- +] | [- -] |
| M | 3;4 | 22 | 4 | 4 | 4 | 4 |
| M | 3;5 | 27 | 4 | 4 | 4 | 4 |
| M | 3;6 | 54 | 4 | 4 | 4 | 4 |
| F | 3;6 | 45 | 4 | 4 | 4 | 3 |
| F | 3;8 | 39 | 4 | 4 | 4 | 4 |
| M | 3;9 | 50 | 4 | 4 | 4 | 4 |
| M | 3;9 | 45 | 4 | 4 | 4 | 4 |
| F | 3;9 | 59 | 4 | 4 | 3/3 | 4 |
| M | 3;9 | 98 | 4 | 4 | 4 | 4 |
| F | 3;10 | 47 | 4 | 4 | 4 | 4 |
| F | 3;10 | 61 | 4 | 4 | 4 | 4 |
| F | 3;10 | 61 | 4 | 4 | 4 | 4 |
| M | 3;11 | 67 | 4 | 4 | 4 | 4 |

*Out of 4 unless otherwise noted

**Table C.7 Individual production scores per sentence type: Experiment 2b (2½ yrs)**

| Gender | Age | Number of SO order productions* | | | |
|---|---|---|---|---|---|
| | | [++] | [+ -] | [- +] | [- -] |
| F | 2;8 | 0/2 | 2/3 | 2/3 | 1/3 |
| M | 2;9 | 2/2 | 2/2 | 2/3 | 2/2 |
| M | 2;9 | 3/3 | 2/4 | 1/4 | 2/3 |
| F | 2;9 | 4 | 4 | 4 | 4 |
| M | 2;10 | 2/2 | 2/2 | 2/2 | 3/3 |
| F | 2;11 | 1/3 | 4 | 3/3 | 1/2 |
| M | 2;11 | 2/2 | 2/2 | 2/2 | 3/3 |
| M | 2;11 | 3/3 | 3 | 4 | 3 |
| F | 3;0 | 2 | 4 | 2 | 4 |
| F | 3;2 | 2/2 | 3/4 | 1/2 | 2/2 |

*Out of 4 unless otherwise noted

**Table C.8 Individual production scores per sentence type: Experiment 2b (3½ yrs)**

| Gender | Age | Number of SO order productions* | | | |
|---|---|---|---|---|---|
| | | [++] | [+ -] | [- +] | [- -] |
| M | 3;3 | 2 | 4 | 4 | 4 |
| F | 3;3 | 3/3 | 4 | 4 | 4 |
| M | 3;3 | 3/3 | 2 | 4 | 3 |
| M | 3;3 | 3/3 | 4 | 3/3 | 3/3 |
| F | 3;3 | 4 | 4 | 4 | 4 |
| M | 3;3 | 4 | 3 | 4 | 4 |
| F | 3;3 | 4 | 4 | 4 | 4 |
| F | 3;4 | 2/2 | 3/3 | 4 | 3 |
| F | 3;4 | 3 | 4 | 3/3 | 3/3 |
| M | 3;5 | 2/2 | 4 | 4 | 4 |
| M | 3;5 | 3/3 | 3/3 | 4 | 3/3 |
| F | 3;5 | 4 | 4 | 4 | 4 |
| M | 3;5 | 4 | 4 | 4 | 3/3 |
| M | 3;6 | 1/2 | 4 | 4 | 4 |
| F | 3;6 | 2/2 | 2/2 | 2/2 | 1/2 |
| F | 3;7 | 2/2 | 2/2 | 1/2 | 3/3 |
| M | 3;7 | 4 | 4 | 4 | 3/3 |
| M | 3;8 | 3 | 3 | 4 | 3 |
| M | 3;8 | 4 | 4 | 4 | 4 |
| F | 3;8 | 3 | 4 | 4 | 4 |
| F | 3;9 | 1/2 | 4 | 4 | 4 |
| F | 3;9 | 4 | 4 | 4 | 4 |
| F | 3;10 | 4 | 4 | 4 | 4 |

*Out of 4 unless otherwise noted

# Experiments 4a & 6a (Chapter 6)

**Table C.9 Individual comprehension scores per sentence type: Experiment 4a (Dutch 2½ yrs)**

| Gender | Age | N-CDI percentile | Number of SO interpretations* | | | |
|---|---|---|---|---|---|---|
| | | | [++] | [+ -] | [- +] | [- -] |
| F | 2;5 | .95 | 1 | 4 | 1 | 4 |
| F | 2;5 | .95 | 2/3 | 2/3 | 2 | 2/3 |
| F | 2;6 | 1.00 | 3 | 3 | 1 | 0 |
| F | 2;8 | .65 | 2 | 2 | 3 | 2 |
| F | 2;8 | .20 | 1 | 2 | 3 | 1/3 |
| F | 2;8 | .20 | 1 | 1 | 1 | 4 |
| F | 2;9 | .70 | 0 | 2 | 2 | 2 |
| M | 2;11 | .70 | 3 | 2 | 2 | 2 |
| M | 2;11 | .20 | 2 | 3/3 | 0 | 3 |
| M | 3;0 | .15 | 0/3 | 3 | 2/3 | 2 |
| M | 3;0 | .75 | 1/3 | 4 | 1 | 3 |
| F | 3;1 | .85 | 1/2 | 3/3 | 4 | 1 |
| F | 3;2 | .85 | 2 | 4 | 2 | 1/3 |

*Out of 4 unless otherwise noted

**Table C.10 Individual comprehension scores per sentence type: Experiment 4a (Dutch 3½ yrs)**

| Gender | Age | KINT Score (%) | Number of SO interpretations* | | | |
|---|---|---|---|---|---|---|
| | | | [++] | [+ -] | [- +] | [- -] |
| F | 3;3 | 90 | 2 | 3 | 1 | 1 |
| F | 3;4 | 57 | 4 | 2 | 2 | 4 |
| F | 3;4 | 32 | 4 | 4 | 4 | 3 |
| F | 3;4 | 67 | 3 | 3 | 4 | 2 |
| F | 3;5 | 15 | 3 | 3 | 1 | 2 |
| F | 3;6 | 15 | 3 | 3 | 2 | 4 |
| M | 3;8 | 45 | 4 | 4 | 2 | 4 |
| M | 3;9 | 29 | 2/3 | 0/2 | 2 | 1/3 |
| M | 3;9 | 57 | 4 | 4 | 1 | 3 |
| M | 3;9 | 39 | 3/3 | 4 | 4 | 3/3 |
| F | 3;9 | 50 | 3 | 3 | 3 | 4 |
| F | 3;9 | 37 | 1 | 3 | 2 | 2 |
| M | 3;10 | 48 | 2 | 4 | 1 | 4 |
| M | 3;11 | 47 | 3 | 3 | 3 | 2 |
| M | 3;11 | 49 | 2 | 3 | 1 | 2 |
| F | 4;0 | 59 | 2 | 4 | 1 | 2 |
| F | 4;1 | 43 | 3 | 4 | 3 | 4 |

*Out of 4 unless otherwise noted

**Table C.11 Individual comprehension scores per sentence type: English-speaking preschoolers (2½ yrs)**

| Gender | Age | Number of SO interpretations* | | | |
|---|---|---|---|---|---|
| | | [++] | [+ -] | [- +] | [- -] |
| F | 2;3 | 3/3 | 1/2 | 1/2 | 2/3 |
| F | 2;3 | 2/3 | 2 | 2 | 1 |
| F | 2;4 | 1 | 1 | 3 | 3 |
| F | 2;6 | 2 | 2 | 1 | 2 |
| M | 2;6 | 2 | 1 | 3 | 1 |
| M | 2;7 | 1/2 | 1/2 | 1/2 | 1/2 |
| F | 2;9 | 1/3 | 2/2 | 1/2 | 3/3 |
| M | 2;9 | 1 | 3 | 2 | 2 |
| M | 2;10 | 3 | 2 | 2 | 2 |
| F | 2;10 | 3 | 2 | 3 | 2 |
| F | 2;11 | 1 | 3 | 3 | 1 |
| M | 3;0 | 4 | 4 | 4 | 4 |
| F | 3;0 | 2 | 3 | 2 | 2 |
| F | 3;0 | 2 | 3 | 4 | 1 |
| F | 3;1 | 4 | 4 | 2 | 2 |
| M | 3;1 | 4 | 3 | 3 | 1 |
| M | 3;1 | 2 | 2 | 2 | 2 |

*Out of 4 unless otherwise noted

**Table C.12 Individual comprehension scores per sentence type: English-speaking preschoolers (3½ yrs)**

| Gender | Age | Number of SO interpretations* | | | |
|---|---|---|---|---|---|
| | | [++] | [+ -] | [- +] | [- -] |
| M | 3;4 | 2/2 | 2/2 | 1/2 | 2/2 |
| M | 3;7 | 2/2 | 2/2 | 2/2 | 1/2 |
| M | 3;8 | 1 | 4 | 3 | 4 |
| M | 3;8 | 3 | 3 | 3 | 3 |
| M | 3;8 | 4 | 4 | 4 | 3 |
| M | 3;9 | 2 | 2/3 | 1/2 | 2/2 |
| F | 3;9 | 4 | 4 | 4 | 4 |
| M | 3;10 | 1 | 3 | 3 | 2 |
| M | 3;10 | 4 | 4 | 4 | 3 |
| F | 3;10 | 3 | 3 | 4 | 2 |
| F | 3;11 | 4 | 3 | 2 | 3 |
| M | 3;11 | 3 | 3 | 4 | 3 |

*Out of 4 unless otherwise noted

# Experiments 4b & 6b (Chapter 7)

**Table C.13 Individual production scores per sentence type: Experiment 4b (Dutch 2½ yrs)**

| Gender | Age | N-CDI percentile | Number of SO order productions* | | | |
|---|---|---|---|---|---|---|
| | | | [++] | [+ -] | [- +] | [- -] |
| F | 2;5 | .95 | 4 | 4 | 2/4 | 2/3 |
| M | 2;11 | .70 | 3/3 | 2 | 2 | 3/3 |
| M | 3;0 | .75 | 4 | 3/3 | 2/3 | 2/3 |
| F | 3;1 | .85 | 2/2 | 4 | 1/3 | 1/3 |
| F | 3;2 | .85 | 4 | 4 | 4 | 4 |

*Out of 4 unless otherwise noted

**Table C.14 Individual production scores per sentence type: Experiment 4b (Dutch 2½ yrs)**

| Gender | Age | KINT score (%) | Number of SO order productions* | | | |
|---|---|---|---|---|---|---|
| | | | [++] | [+ -] | [- +] | [- -] |
| F | 3;4 | 57 | 4 | 4 | 4 | 3 |
| F | 3;4 | 32 | 1/3 | 1/3 | 1/2 | 1/3 |
| F | 3;4 | 67 | 4 | 4 | 4 | 4 |
| F | 3;5 | 15 | 4 | 4 | 4 | 4 |
| F | 3;6 | 15 | 4 | 4 | 4 | 4 |
| M | 3;8 | 45 | 4 | 4 | 4 | 3 |
| M | 3;9 | 29 | 4 | 4 | 4 | 4 |
| M | 3;9 | 57 | 4 | 4 | 3/3 | 4 |
| M | 3;9 | 39 | 3/3 | 4/4 | 3/3 | 3/3 |
| F | 3;9 | 50 | 4 | 4 | 4 | 4 |
| F | 3;9 | 37 | 3 | 4 | 3 | 2 |
| M | 3;10 | 48 | 4 | 4 | 4 | 4 |
| M | 3;11 | 47 | 3 | 4 | 2 | 3 |
| M | 3;11 | 49 | 2/2 | 2/2 | 2/2 | 2/2 |
| F | 4;0 | 59 | 4 | 4 | 4 | 4 |
| F | 4;1 | 43 | 3/3 | 2/2 | 2/2 | 2/2 |

*Out of 4 unless otherwise noted

**Table C.15 Individual production scores per sentence type: Experiment 6b (English-speaking 2½ yrs)**

| Gender | Age | Number of SO order productions* | | | |
|---|---|---|---|---|---|
| | | [++] | [+ -] | [- +] | [- -] |
| F | 2;4 | 1/2 | 2/3 | 0/2 | 2/2 |
| F | 2;4 | 2/2 | 2/2 | 0/2 | 2/2 |
| M | 2;6 | 3/3 | 3/3 | 3/3 | 4 |
| M | 2;9 | 3 | 3 | 3 | 2/3 |
| M | 2;10 | 2/3 | 4 | 4 | 3 |
| F | 2;10 | 3/3 | 2/2 | 1/3 | 2/2 |
| F | 2;11 | 3 | 3 | 2 | 3 |
| M | 3;0 | 4 | 4 | 4 | 4 |
| F | 3;0 | 4 | 2 | 2 | 4 |
| F | 3;0 | 2/2 | 2/2 | 1/2 | 2/2 |
| F | 3;1 | 3/3 | 4 | 3 | 3/3 |
| M | 3;1 | 4 | 4 | 3 | 4 |
| M | 3;1 | 0/3 | 3 | 2 | 0/3 |
| F | 3;2 | 3 | 4 | 1/3 | 2/2 |

*Out of 4 unless otherwise noted

**Table C.16 Individual production scores per sentence type: Experiment 6b (English-speaking 3½ yrs)**

| Gender | Age | Number of SO order productions* | | | |
|---|---|---|---|---|---|
| | | [++] | [+ -] | [- +] | [- -] |
| M | 3;4 | 2/2 | 2/2 | 2/2 | 2/2 |
| M | 3;7 | 2/2 | 2/2 | 2/2 | 2/2 |
| M | 3;8 | 3/3 | 3/3 | 3 | 4 |
| M | 3;8 | 4 | 4 | 4 | 4 |
| M | 3;8 | 4 | 2 | 2 | 4 |
| F | 3;9 | 4 | 4 | 4 | 3/3 |
| M | 3;10 | 3 | 2 | 1 | 4 |
| M | 3;10 | 4 | 4 | 4 | 4 |
| F | 3;10 | 3/3 | 2/2 | 4 | 2/2 |
| F | 3;11 | 4 | 4 | 4 | 4 |
| M | 3;11 | 4 | 4 | 4 | 4 |

*Out of 4 unless otherwise noted

# Bibliography

Agresti, A. (2002). *Categorical data analysis* (1st ed.). New York: Wiley-Interscience.
Aissen, J. (1997). On the syntax of obviation. *Language*, *73*(4), 705-750.
Aissen, J. (2003). Differential object marking: Iconicity vs . economy. *Natural Language and Linguistic Theory*, *21*(3), 435-483.
Aissen, J. (2004). Differential coding, partial blocking, and bidirectional OT. In P. Nowak & C. Yoquelet (Eds.), *Berkeley Linguistics Society 29* (pp. 1-16). Berkeley Linguistics Society.
Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition, Vol. 1: The data* (Vol. 33, pp. 3-24). Hillsdale, N.J.: Lawrence Erlbaum.
Allan, K. (2008). Hierarchies and the choice of left conjuncts (with particular attention to English). *Journal of Linguistics*, *23*(01), 51-77.
Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419-439.
Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247-264.
Angiolillo, C. J., & Goldin-Meadow, S. (1982). Experimental evidence for agent-patient categories in child language. *Journal of Child Language*, *9*, 627-643.
Artstein, R. (2000). Case constraints and empty categories in Optimality Theory parsing.
Baayen, H. R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R. Sociolinguistic Studies* (Vol. 2).
Baayen, H. R., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. Elsevier Inc.
Bates, D. M. (2007). *lme4: Linear mixed-effects models using S4 classes* (Vol. R package).
Bates, E., & Devescovi, A. (1989). Crosslinguistic studies of sentence processing. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 225-253). Cambridge: Cambridge University Press.
Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of art* (pp. 173-217). Cambridge: Cambridge University Press.
Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157-193). Hillsdale, New Jersey: Laurence Erlbaum Associates.
Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 3-73). Cambridge: Cambridge University Press.
Bates, E., Dale, P. S., & Thal, D. J. (1995). Individual differences and their implications for theories of lanugage development. In P. Fletcher & B. MacWhinney (Eds.), *Methodology*. Oxford: Basil Blackwell.
Bates, E., MacWhinney, B., Caselli, C., Devescovi, A., Natale, F., & Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child Development*, *55*(2), 341-354.
Bavin, E. L., & Shopen, T. (1989). Cues to sentence interpretation in Warlpiri. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 185- 205). Cambridge, MA.
van Bergen, G. (2011). *Who's first and what's next: Animacy and word order variation in Dutch language production*. Nijmegen: Radboud University Nijmegen.
Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the Development of Language* (pp. 279–362). New York: Wiley.
Bever, T., Mehler, J., & Valian, V. (1970). *Linguistic capacity of very young children* (pp. 1-9).

Bloom, P., Barss, A., Nicol, J., & Conway, L. (1994). Children's knowledge of binding and coreference: Evidence from spontaneous speech. *Language*, *70*(1), 53-71.

Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics, 17*, 189-216.

Bock, J. K., & Loebell, H. (1990). Framing sentences. *Cognition, 35*, 1-39.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*(1), 47-67.

Bock, J. K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world paradigm* (pp. 249-278). Psychology Press.

Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, *99*(1), 150-171.

Boersma, P. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Science*, *21*, 43-58.

Boersma, P., & Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry, 32*(1), 45-86.

Boersma, P., & Levelt, C. C. (1999). Gradual constraint-ranking learning algorithm predicts acquisition order. *Proceedings of 30th Child Language Research Forum*. Stanford.

Bouma, G. J. (2008). *Starting a sentence in Dutch: a corpus study of subject- and object-fronting*. Groningen: University of Groningen.

Braine, M. D. S., & Bowerman, M. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development*, *41*(1), 1-104.

Brandt-Kobele, O.C., & Höhle, B. (2010). What asymmetries within comprehension reveal about asymmetries between comprehension and production : The case of verb inflection in language acquisition. *Lingua*, *120*, 1910-1925.

Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua, 118*(2), 172-189.

Bresnan, J. (2000). Optimal syntax. In J. Dekkers, F. van Der Leeuw, & J. van De Weijer (Eds.), *Theory: Phonology, syntax and acquisition* (pp. 334-385). Oxford: Oxford University Press.

Bresnan, J., & Aissen, J. (2002). Optimality and funtionality: Objections and refutations. *Natural Language and Linguistic Theory*, *20*(1), 81-95.

Brown, R. (1973). *A first language: The early stages*. Cambridge MA: Harvard University Press.

Chan, A., Lieven, E., & Tomasello, M. (2009). Children's understanding of the agent-patient relations in the transitive construction: Cross-linguistic comparisons between Cantonese, German, and English. *Cognitive Linguistics*, *20*(2), 267-300.

Chan, A., Meints, K., Lieven, E., & Tomasello, M. (2010). Young children's comprehension of English SVO word order revisited: Testing the same children in act-out and intermodal preferential looking tasks. *Cognitive Development*, *25*(1), 30-45.

Chapman, R. S., & Kohn, L. L. (1978). Comprehension strategies in two and three year olds: Animate agents or probable events? *Journal of Speech and Hearing Research*, *21*, 746-761.

Chapman, R. S., & Miller, J. F. (1975). Word order in early two and three word utterances: Does production precede comprehension? *Journal of Speech and Hearing Research*, *18*, 355-371.

Chien, Y., & Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition, 1*(3), 225-295.

Childers, J. B., & Tomasello, M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology, 37*(6), 739-748.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N., & Lasnik, H. (1993). Principles and Parameters Theory. In J. Jacobs, A. von Stechow, W. Sternefeld, & T. Vennemann (Eds.), (pp. 506-569). Berlin: Walter de Gruyter.

Clahsen, H. (1982). *Spracherwerb in der Kindheit. Eine Untersuchung zur Entwicklung der Syntax bei Kleinkinderen*. Tuebingen: Gunter Narr Verlag.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(01), 3-42.

Clark, E. V., & Hecht, B. F. (1983). Comprehension, production, and language acquisition. *Annual Review of Psychology*, *34*(1), 325-349.

Clark, H. H. (1965). Some structural properties of simple active and passive sentences. *Journal of Verbal Learning and Verbal Behavior*, *4*, 365-370.

Clifton, C., Traxler, M. J., Mohamed, T. M., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*(3), 317-334.

Cocking, R., & McHale, S. (1977). A comparative study of the use of pictures and objects in assessing children's receptive and productive language. *Annual Meeting of the American Psychological Association 85*. San Francisco.

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review Psychology*, *52*, 337-67.

Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology* (2nd ed.). Oxford: Basil Blackwell, Inc.

Conroy, A., Takahashi, E., Lidz, J., & Phillips, C. (2009). Equal treatment for all antecedents: How children succeed with Principle B. *Linguistic Inquiry*, *40*(3), 446-486.

Corrigan, R. (1986). The internal structure of English transitive sentences. *Memory and Cognition*, *14*(5), 420-31.

Corrigan, R., & Odya-Weis, C. (1985). The comprehension of semantic relations by two-year-olds: An exploratory study. *Journal of Child Language*, *12*, 47-59.

Crain, S., & Thornton, R. (1998). *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.

Crain, S., Goro, T., & Minai, U. (2007). Hidden units in child language. In Schalley & Khlentzos (Eds.), *Mental States Volume 1: Evolution; Function; Nature*. Amsterdam: John Benjamins Publishers.

Crain, S., Thornton, R., Boster, C., Conway, L., Lillo-Martin, D., & Woodams, E. (1996). Quantification without qualification. *Language Acquisition*, *5*(2), 83-153.

Dahl, Ö. (2000). Egophoricity in discourse and syntax. *Functions of Language*, *7*, 33-77.

Dahl, Ö., & Fraurud, K. (1996). Animacy in grammar and discourse. In T. Fretheim & J. K. Gundel (Eds.), *Reference and referent accessibility* (pp. 47-64). Amsterdam/Philadelphia: John Benjamins.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior research methods, instruments, & computers*, *28*, 125-127.

Davies, C., & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua*, *120*(8), 1956-1972.

Demuth, K. (1996). Collecting spontaneous production data. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 3-22). Cambridge, MA: MIT Press.

Dewart, H. M. (1979). The role of animate and inanimate nouns in determining sentence voice. *British Journal of Psychology*, *70*, 135-141.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development*, *79*(4), 1152-67.

Drenhaus, H., & Fery, C. (2008). Animacy and child grammar : An OT account. *Lingua*, *118*, 222-244.

Dryer, M. S. (2008). Order of subject, object and verb. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures online*. Munich. Retrieved from http://wals.info/feature/81

Déprez, V., & Pierce, A. (1993). Negation and functional projections in early grammar. *Linguistic Inquiry*, *24*, 25-67.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409-436.

Eisenbeiss, S. (2010). Production methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 11-34). Amsterdam/Philadelphia: John Bemjamins Publishing Company.

Fanselow, G., Schlesewsky, M., Cavar, D., & Kliegl, R. (1999). Optimal parsing: Syntactic parsing preferences and Optimality Theory.

Feldman, H. (1986). *A Grammar of Awtuw*. Canberra, Australia: Research School of Pacific Studies, The Australian National University.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. . (2000). Short form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics, 21*, 95-115.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-Line methods in children's language processing* (pp. 184-218). Amsterdam: John Benjamins.

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, *33*, 715-736.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*(2), 164-203.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11-15.

Fikkert, P., & Levelt, C. C. (2004). How does place fall into place? The lexicon and emergent constraints in the developing phonological grammar. In P. Avery, B. E. Dresher, & K. Rice (Eds.), *Contrast in phonology: Perception and acquisition*. Berlin: Mouton.

Fikkert, P., & de Hoop, H. (2009). Language acquisition in Optimality Theory. *Linguistics*, *47*(2), 311-357.

Ford, M., Bresnan, J. W., & Kaplan, R. M. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 727-796). Cambridge, MA: MIT Press.

Fraser, C., Bellugi, U., & Brown, R. (1963). Control of grammar in imitation, comprehension, and production. *Journal of Verbal Learning and Verbal Behavior, 2*, 121-135.

Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, *5*, 519-559.

Frazier, L., & Flores D'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, *28*(3), 331-344.

Friedman, S., & Stevenson, M. (1975). Developmental changes in the understanding of implied motion in two-dimentional pictures. *Child Development*, *46*, 773-778.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Gerken, L., & Shady, M. E. (1996). The picture selection task. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 125-145). Cambridge, MA: MIT Press.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, *17*(8), 684-691.

Gibson, E., & Broihier, K. (1993). Optimality theory and human sentence processing. *English*, (Dir 9113607).

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and language*, *57*(4), 544-569.

Gnanadesikan, A. E. (2004). Markedness and faithfulness constraints in child phonology. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in phonological acquisition* (pp. 73-108). Cambridge, MA: Cambridge University Press.

Goldin-Meadow, S., Seligman, M. E. P., & Gelman, R. (1976). Language in the two-year old. *Cognition*, *4*, 189-202.

Goodluck, H. (1996). The act-out task. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 147-162). Cambridge, MA: MIT Press.

Gough, P. B. (1966). The verification of sentences: The effects of delay of evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior, 5*, 492-496.

Griffin, Z. M., & Bock, J. K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 247-279.

Grimshaw, J. (1995). Projection, heads, and optimality. *English*. New Brunswick, New Jersey.

Grimshaw, J., & Rosen, S. T. (1990). Knowledge and obedience: The developmental status of the binding theory. *Linguistic Inquiry*, *21*(2), 187-222.

Gualmini, A. (2004). Some knowledge children don't lack. *Linguistics*, *42*(5), 957-982.

Guasti, M. (2002). *Language acquisition: The growth of grammar*. Cambridge, MA: The MIT Press.

Hakuta, K. (1982). Interaction between particles and word order in the comprehension and production of simple sentences in Japanese children. *Developmental Psychology*, *18*(1), 62-76.

Hamann, C., Kowalski, O., & Philip, W. (1997). The French "Delay of Principle B" effect. In E. Hughes, M. Hughes, & A. Greenhil (Eds.), *BUCLD 21: Proceedings of the 21st annual Boston University Conference on Language Development* (pp. 205-219). Somerville, Mass: Cascadilla Press.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, *15*(6), 409-414.

Hendriks, P. (2008). A unified explanation for production/comprehension asymmetries. In A. G. Algueró & M. João Freitas (Eds.), *Proceedings of GALA 2007* (pp. 240-251). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, *120*(8), 1887-1897.

Hendriks, P., & Spenader, J. (2004). A bidirectional explanation of the pronoun interpretation problem. In P. Schlenker & E. Keenan (Eds.), *Proceedings of the ESSLLI '04 Workshop on Semantic Approaches to Binding Theory*. Nancy, France.

Hendriks, P., & Spenader, J. (2005/2006). When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition*, *13*(4), 319-348.

Hendriks, P., & de Hoop, H. (2001). Optimality theoretic semantics. *Linguistics and Philosophy*, *24*, 1-32.

Hendriks, P., de Hoop, H., & Lamers, M. J. A. (2005). Asymmetries in language use reveal asymmetries in the grammar. In P. Dekker & M. Franke (Eds.), *Proceedings of the 15th Amsterdam Colloquium, ILLC* (pp. 113-118). Amsterdam.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996a). *The origins of grammar: Evidence from early language comprehension*. (D. McDaniel, C. McKee, & H. S. Cairns, Eds.)*Methods for assessing children's syntax* (pp. 105-124). Cambridge, MA: The MIT Press.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996b). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 105-124). Cambridge, MA: MIT Press.

Hoeks, J. C. J., & Hendriks, P. (2005). Optimality Theory and human sentence processing: The case of coordination. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 959-964). Mahwah, NJ: Erlbaum.

Hogeweg, L., & de Hoop, H. (2010). Children and transitivity: The subject-object asymmetry in a natural setting. In P. Brandt & M. Garcia Garcia (Eds.), *Transitivity; form meaning and processing* (pp. 143 - 160). Amsterdam: John Bemjamins Publishing Company.

de Hoop, H., & Krämer, I. (2005/6). Children's optimal interpretations of indefinite subjects and objects. *Language Acquisition*, *13*(2), 103-123.

de Hoop, H., & Lamers, M. J. A. (2006). Incremental distinguishability of subject and object. In L. Kulikov, A. Malchukov, & P. de Swart (Eds.), *Case, valency and transitivity*. Amsterdam: John Benjamins.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One frog , two frog , red frog , blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research*, *29*(6), 597-626.

Ingram, D. (1974). The relationship between comprehension and production. In R. Schifelbusch & L. Lloyd (Eds.), *Language perspectives: Acquisition, retardation and intervention* (pp. 313-334). Baltimore: University Park Press.

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of memory and language*, *59*(4), 434-446.

Johnson, E., & Zamuner, T. (2010). Using infant and toddler testing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 73-93). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Jäger, G. (2003). Learning constraint sub-hierarchies. The bidirectional Gradual Learning Algorithm. In R. Blutner & H. Zeevat (Eds.), *Pragmatics in OT* (pp. 251-287). Houndmills: Palgrave MacMillan.

Kaan, E. (1997). *Processing subject-object ambiguities in Dutch*. Groningen: University of Groningen.

Kager, Rene. (1999). *Optimality Theory*. Cambridge: Cambridge University Press.

Kail, M. (1989). Cue validity, cue cost, and processing types in sentence comprehension in French and Spanish. In B MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 77 - 117). Cambridge, MA: Cambridge University Press.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*, 113-147.

Keenan, J. M., & Macwhinney, B. (1987). Understanding the relationship between comprehension and production. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 149-155). Norwood, NJ: Ablex Publishing Corporation.

Kempen, G., & Harbusch, K. (2004). A corpus study into word order variation in German subordinate clauses: Animacy affects linearization independently of grammatical function assignment. In T. Pechmann & C. Habel (Eds.), *Multidisciplinary approaches to language production* (pp. 173-181). Berlin: Mouton De Gruyter.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201-258.

Knoeferle, P., & Crocker, M. W. (2005). Incremental effects of mismatch during picture-sentence integration. In Bruno G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Society* (pp. 1166-1171). Mahwah, NJ: Lawrence Erlbaum.

Koff, E., Kramer, P. E., & Fowles, B. (1980). Effects of event probability and animateness on children's comprehension of active and passive sentences. *The Journal of Psychology, 104*, 157-163.

Koster, C. (1993). Errors in anaphora acquisition. Utrecht: Utrecht University.

Koster, C., Plas, J. J., & Krikhaar, E. (2004). Handleiding KINT (Kleuter Inventarisatie Nederlandse Taalverwerving). Groningen.

Krämer, I. (2000). *Interpreting indefinites: An experimental study of children's language comprehension*. Nijmegen: MPI.

Kuhn, J. (2000). Generation and parsing in Optimality Theoretic syntax: Issues in the formalization of OT-LFG. In P. Sells (Ed.), *Formal and empirical issues in Optimality Theoretic syntax*. Stanford, CA: CSLI Publications.

Kuijper, S. J. M., & Groothoff, F. C. (2010). Children's comprehension and production of marked stress. In K. Franich, K. M. Iserman, & L. L. Keil (Eds.), *BUCLD 34: Proceedings of the 34rd annual Boston University Conference on Language Development* (pp. 233-244). Somerville, MA: Cascadilla Press.

Lamers, M. J. A. (2005). The on-line resolution of subject-object ambiguities with and without case-marking in Dutch: Evidence from event-related brain potentials. In M. Amberber & H. de Hoop (Eds.), *Competition and variation in natural languages: The case for case*. Oxford: Elsevier Ltd.

Lamers, M. J. A., & de Hoop, H. (2004). The role of animacy information in human sentence processing captured in four conflicting constraints. In H. Christiansen, P. R. Skadhauge, & J. Villadsen (Eds.), *Preliminary Proceedings of the International Workshop on Constraint Solving and Language Processing (CSLP 2004)* (pp. 102-113). Copenhagen/Roskilde: Roskilde Department of Computer Science, Roskilde University.

Lamers, M. J. A., & de Hoop, H. (2005). Animacy information in human sentence processing: An incremental optimization of interpretation approach. In H. Christiansen, P. R. Skadhauge, & J. Villadsen (Eds.), *Constraint Solving and Language Processing (Lecture Notes in Artificial Intelligence 3438)* (pp. 158-171). Berlin/Heidelberg: Springer- Verlag.

Legendre, G. (2001). An introduction to Optimality Theory in syntax. In P. Sells (Ed.), *Optimality-theoretic syntax* (pp. 1-27). Stanford, CA: MIT Press.

Legendre, G., Hagstrom, P., Vainikka, A., & Todorova, M. (2006). Optimality in language acquisition II: Inflection in early French syntax. In P. Smolensky & G. Legendre (Eds.), *The Harmonic Mind: From neural computation to Optimality-Theoretic grammar, vol. 2* (Vol. 2, pp. 279-306). Cambridge: MIT Press.

Lempert, H. (1989). Animacy constraints on preschool children's acquisition of syntax. *Child Development*, *60*(1), 237-245.

Lempert, H. (1990). Acquisition of passives: The role of patient animacy, salience, and lexical accessibility. *Journal of Child Language*, *17*, 677-696.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-75.

Lieven, E., & Ambridge, B. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge: Cambridge University Press.

Lindner, K. (2003). The development of sentence-interpretation strategies in monolingual German-learning children with and without specific language impairment. *Linguistics*, *41*(2), 213-254.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Laurence Erlbaum Associates.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 127-150.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, *47*(1), 50-68.

McClellan, J., Yewchuk, C., & Holdgrafer, G. (1986). Comprehension and production of word order by 2-year-old children. *Journal of Psycholinguistic Research*, *15*(2), 97-116.

McDaniel, D., McKee, C., & Cairns, H. S. (1996). *Methods for assessing children's syntax*. Cambridge, MA: MIT Press.

McDonald, J. L. (1987). Assigning linguistic role: The influence of conflicting cues. *Journal of Memory and Language*, *26*, 100-117.

McDonald, J. L., Bock, J. K., & Kelly, M. H. (1993). Word and world order: semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology, 25*, 188-230.

McKee, C. (1992). A comparison of pronouns and anaphors in Italian and English acquisition. *Language Acquisition*, *2*(1), 21-54.

McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy*, *6*(2), 203-229.

Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory and Cognition*, *23*(4), 477-94.

Meisel, J. (1995). Parameters in acquisition. In B. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 10-35). Boston: Blackwell.

Meroni, L., & Crain, S. (2002). On not being led down the kindergarten path. Proceedings of the 27th Boston University Conference on Language Development, 531-544.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. Psychological Review, 104(4), 686-713.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. Psychological Science, 13(4), 329-336.

Naigles, L. R. (1990). Children use syntax to learn verb meanings. Journal of Child Language, 17, 357-374.

Neeleman, A., & Weerman, F. (1997). L1 and L2 word order acquisition. Language Acquisition, 6(2), 125-170.

Newmeyer, F. J. (2002). Optimality and functionality: A critique of functionally-based Optimality-Theoretic syntax. *Natural Language and Linguistic Theory*, *20*(1), 43-80.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785-806.

Øvrelid, L. (2004). Disambiguation of syntactic functions in Norwegian: Modeling variation in word order interpretations conditioned by animacy and definiteness. In F. Karlsson (Ed.), *Proceedings of the 20th Scandinavian Conference of Linguistics*. Helsinki: University of Helsinki.
Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition*, *86*, 253-282.
Papafragou, A., Hulbert, J., & Trueswell, J. C. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*, 155-184.
Philip, W., & Coopmans, P. (1996). The double Dutch delay of Principle B effect. In A. Stringfellow, E. Cahana-Amitay, E. Hughes, & A. Zukowski (Eds.), *The Proceedings of the 20th annual Boston University Conference on Language Development* (pp. 576-587). Somerville, MA: Cascadilla Press.
Phillips, C. (1995). Right association in parsing and grammar. In C. Schütz, K. Broihier, & J. Ganger (Eds.), *Papers on Language Processing and Acquisition* (pp. 37-93). Cambridge, MA: MIT Press.
Platzack, C. (1996). The initial hypothesis of syntax: A minimalist perspective on language acquisition and attrition. In Harald Clahsen (Ed.), *Generative perspectives on language acquisition: Empirical findings, theoretical considerations and cross-linguistic comparison* (pp. 161-200). Amsterdam: John Benjamins.
Pléh, C. (1989). The development of sentence interpretation in Hungarian. In B. Macwhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing*. Cambridge: Cambridge University Press.
Poeppel, D., & Wexler, K. (1993). The full competence hypothesis of clause structure in early German. *Language*, *69*(1), 1-33.
Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: a cross-linguistic study in English & Spanish. *Journal of Memory and Language*, *42*, 168-182.
Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar*. Boulder: Blackwell.
R Development Core Team. (2011). *R: A language and environment forstatistical computing*. Vienna, Austria.
Ramer, A. (1976). Syntactic styles in emerging language. *Journal of Child Language*, *3*, 49-62.
Reinhart, T. (2004). The processing cost of reference-set computation: Acquisition of stress shift and focus. *Language Acquisition*, *12*(2), 109-155.
Reinhart, T. (2006). *Interface strategies: Optimal and costly computations*. Cambridge, MA: MIT Press.
Samek-Lodovici, V. (2007). An invitation to OT syntax and semantics. *Lingua*, *117*, 1513-1521.
Scerif, G., Karmiloff-Smith, A., Campos, R., Elsabbagh, M., Driver, J., & Cornish, K. (2005). To look or not to look? Typical and atypical development of oculomotor control. *Journal of cognitive neuroscience*, *17*(4), 591-604.
Schaeffer, J. (2000). *The acquisition of direct object scrambling and clitic placement: Syntax and pragmatics*. (S. M. Powers & C. Hamann, Eds.)*The acquisition of scrambling and cliticization* (pp. 71-93). Amsterdam: John Benjamins.
Schaner-Wolles, C. (1989). Strategies in acquiring grammatical relations in German: Word order or case marking? *Folia Linguistica*, *23*(1), 131-156.
Schlesewsky, M., & Bornkessel, I. (2004). On incremental interpretation: Degrees of meaning accessed during sentence comprehension. *Lingua*, *114*, 1213-1234.
Schmitt, C., & Miller, K. (2010). Using comprehension methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 35-56). Amsterdam/Philadelphia: John Bemjamins Publishing Company.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide*. (L. R. and D. Center, Ed.). University of Pittsburgh: Psychology Software Tools, Inc.
Schriefers, H., Friederici, A. D., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, *34*, 499-520.
Scott, G. (1987). *The Fore language of Papa New Guinea*. Canberra, Australia: Canberra Research School of Pacific Studies, The Australian National University.

Sedivy, J. C. (2010). Using eyetracking in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 115-139). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Sekerina, I. A., Stromswold, K., & Hestvik, A. (2004). How do adults and children process referentially ambiguous pronouns? *Journal of Child Language*, *31*, 123-152.

Shatz, M. (1978). On the development of communicative understandings: An early strategy for interpreting and responding to messages. *Cognitive Psychology*, *10*, 271-301.

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112-171). Canberra, Australia: Australian Institute of Aboriginal Studies.

Sinclair, H., & Bronckart, J. P. (1972). S.V.O A linguistic universal? A study in developmental psycholinguistics. *Journal of Experimental Child Psychology*, *14*, 329-348.

Slobin, D. I. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, *5*, 219-227.

Slobin, D. I., & Bever, T. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, *12*, 229-265.

de Smedt, K. J. M. J. (1990). IPF: An incremental parallel formulator. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation, Cognitive Science Series* (pp. 167–192). San Diego: Academic Press.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry*, *27*, 720-731.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238-99.

Snider, N., & Zaenen, A. (2006). Animacy and syntactic structure: Fronted NPs in English. In M. Butt, M. Dalrymple, & T. H. King (Eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan* (pp. 323-338). CSLI Publications.

Sokolov, J. L. (1989). The development of role assignment in Hebrew. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 118-157). Cambridge: Cambridge University Press.

Spenader, J., Smits, E.-J., & Hendriks, P. (2009). Coherent discourse solves the pronoun interpretation problem. *Journal of Child Language*, *36*(1), 23-52.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*, 447-481.

Stevenson, S., & Smolensky, P. (2006). Optimality in sentence processing. In P. Smolensky & G. Legendre (Eds.), *The harmonic mind: From neural computation to Optimality-Theoretic grammar, vol. 2* (Vol. 2, pp. 307-338). London: MIT Press.

Strohner, H., & Nelson, K. E. (1974). The young child's development of sentence comprehension: Influence of event probability, nonverbal context, syntactic form, and strategies. *Child Development*, *45*(3), 567-576.

Stromswold, K. (1996). Analyzing children's spontaneous speech. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 23-53). Cambridge, MA: MIT Press.

de Swart, Peter. (2007). *Cross-linguistic variation in object marking*. Nijmegen: Radboud University Nijmegen.

de Swart, Peter, & Cannizzaro, G. (2012). The use of animacy information in online sentence processing: Evidence from anticipatory eye-movements. Groningen/Nijmegen manuscript.

Tanaka, M., Branigan, H. P., & Pickering, M. J. (2005). The role of animacy in Japanese sentence production. *Paper Presented at the CUNY Sentence Processing Conference*. Tucson, AZ.

Tesar, B., & Smolensky, P. (1998). Learnability in Optimality Theory. *Linguistic Inquiry*, *29*(2), 229-268.

Thal, D. J., & Flores, M. (2001). Development of sentence interpretation strategies by typically developing and late-talking toddlers. *Journal of Child Language*, *28*, 173-193.

Thornton, R. (1996). Elicited production. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 77-102). Cambridge, MA: MIT Press.
Thornton, R. (2002). Let's change the subject: Focus movement in early grammar. *Language Acquisition*, *10*(3), 229-271.
Thornton, R., & Wexler, K. (1999). *Principle B, VP ellipsis and interpretation in child grammar*. Cambridge, MA: MIT Press.
Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press.
Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, *74*, 209-253.
Tomlin, R. (1986). The animated first principle. *Basic word order. Functional principles* (pp. 102-139). London: Croom Helm.
Tomlin, R. (1995). Focal attention, voice, and word order. In P. Downing & M. Noonan (Eds.), *Word order in discourse* (pp. 517-554). Amsterdam: John Benjamins.
Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*, 69-90.
Trueswell, J. C. (2008). Using eye movements as a developmental measure within psycholinguistics. *Developmental psycholinguistics: On-line methods in children's language processing* (Vol. i, pp. 139-183).
Trueswell, J. C., & Tanenhaus, M. K. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285-318.
Trueswell, J. C., Sekerina, I. A., Hill, N. M., & Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89-134.
Turner, E. A., & Rommetveit, R. (1967). The acquisition of sentence voice and reversibility. *Child Development*, *38*(3), 649-660.
Unsworth, S. (2007). L1 and L2 acquisition between sentence and discourse: Comparing production and comprehension in child Dutch. *Lingua*, *117*, 1930-1958.
Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, *40*(1-2), 21-81.
Van Nice, K. Y., & Dietrich, R. (2003). Task sensitivity of animacy effects: Evidence from German picture descriptions. *Linguistics*, *41*(5), 825-849.
Van Valin, J. R. D., & Wilkins, D. P. (1996). The case for "effector": Case roles, agents, and agency revisited. In M. Shibatani & S. A. Thompson (Eds.), *Grammatical constructions* (pp. 289-322). Oxford: Oxford University Press.
Verrips, M., & Weissenborn, J. (1992). Routes to verb placement in early German and French: The independence of finiteness and agreement. In J. Meisel (Ed.), *The acquisition of verb placement: Functional categories and V2 phenomena in language acquisition* (pp. 283-331). Dordrecht: Kluwer.
de Villiers, J. G., & de Villiers, P. A. (1973). Development of the use of word order in comprehension. *Journal of Psycholinguistic Research*, *2*(4), 331-341.
de Villiers, J. G., Cahillane, J., & Altreuter, E. (2006). What can production reveal about Principle B? In K. U. Deen, J. Nomura, B. Schul, & B. D. Schwartz (Eds.), *The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition–North America* (4th ed., pp. 89-100). University of Connecticut Occasional Papers in Linguistics.
Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*, B63-B72.
Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, *36*, 559 - 570.
Weinberg, A. (1993). Parameters in the theory of sentence processing: Minimal Commitment Theory goes east. *Journal of Psycholinguistic Research*, *22*(3), 339-364.
Weinberg, A. (2001). A minimalist theory of human sentence processing. In Epstein & Hornstein (Eds.), *Working minimalism*. Cambridge, MA: MIT Press.

Weissenborn, J. (1992). Null subjects in early grammars: Implications for parameter-setting theories. In J. Weissenborn, H. Goodluck, & T. Roeper (Eds.), *Theoretical issues in language acquisition: Continuity and change in development* (pp. 269-300). Hillsdale, N.J.: Lawrence Erlbaum.

Wijnen, F., & Verrips, M. (1998). The acquisition of Dutch syntax. In S. Gillis & A. de Houwer (Eds.), (pp. 223-299). Amsterdam: John Benjamins Publishing company.

Wilkins, D. P. (1990). Notions relevant to the discussion of "agents" and "agency" as coded in natural languages. Davis: University of California.

Wubs, E., Hendriks, P., Hoeks, J. C. J., & Koster, C. (2009). Tell me a story! Children's capacity for topic shift. In J. Crawford, K. Otaki, & T. Masahiko (Eds.), *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)* (pp. 313-324). Somerville, MA: Cascadilla Proceedings Project.

Yang, C. (2010). *Who's afraid of George Kingsley Zipf?* Philadelphia: University of Pennsylvania.

Yang, Q., Bucci, M. P., & Kapoula, Z. (2002). The latency of saccades, vergence, and combined eye movements in children and adults. *Investigative Ophthamology and Visual Science*, *43*, 2939-2949.

Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., O'Connor, M. C., et al. (1976). Animacy encoding in English: Why and how. In B. Webber & D. Byron (Eds.), *Proceedings of the 2004 ACL Workshop on Discourse Annotation*. Stroudsburg, PA: Association for Computational Linguistics.

Zink, I., & Lejaegere, M. (2002). *N-CDIs: Lijsten voor communicatieve ontwikkeling. Aanpassing en hernormering van de MacArthur CDIs van Fenson et al., 1993*. Leuven/Leusden: Acco.

Zink, I., & Lejaegere, M. (2003a). *N-CDI's: Kortevormen; Dutch Adaptation And Norms Of MacArthur Short From Vocabulary Checklists Van Fenson et al. 1993*. Leuven: Acco.

Zink, I., & Lejaegere, M. (2003b). *N-CDI 3 Dutch adaptation with norms of the MacArthur CDI: Level III van Dale et al. (2000)*. Leuven: Acco.

# Summary in English

## Chapter 1: The phenomenon

When presented with a sentence like *The car is pushing the cow* and asked to act it out with toys, a child between the age of two-and-a-half and three-and-a-half will be about 50% likely to make the cow push the car (Chan, Lieven, & Tomasello, 2009; Chapman & Miller, 1975). Thus, even without the influence of verb argument requirements or event likelihood, English-speaking children appear to regularly ignore word order in favor of interpretations that allow animate entities to act upon inanimate entities.

This phenomenon in which preschoolers fail to reliably use word order when interpreting reversible transitive sentences is particularly intriguing when set against the backdrop of their apparent adult-like production of word order (Angiolillo & Goldin-Meadow, 1982; Chapman & Miller, 1975; McClellan, Yewchuk, & Holdgrafer, 1986). This is striking since it is more common that comprehension exceeds production. The remarkable possibility that children employ word order in a reversible sentence like the one above differently in comprehension than they do in production inspired the two central themes of this dissertation: asymmetry and animacy. Do children produce subject-object word order more successfully than they are able to understand it? And if so, is their comprehension systematically affected by the relative animacy of subject and object?

## Chapters 2 & 3: Theoretical framework, methodology, and predictions

In order to answer these questions, I first establish in Chapter 2 which theoretical framework that is best able to account for asymmetric performance by young children. Frameworks unable to account for both variable comprehension and adult-like production were ruled out in favor of Optimality Theory, which treats comprehension and production as two different “directions” of the same grammar. The nature of the hierarchical system of soft constraints, some of which promote symmetry and others which promote asymmetry, allows for a comprehensive account of the early word order asymmetry.

Two theoretical models within the Optimality Theory framework were utilized to make specific predictions for early word order, as well as for adult word order, in Dutch and English. The first model, of Hendriks, de Hoop, and Lamers (2005), allowed for offline predictions. Their proposal involves an animacy constraint that is incorrectly ranked above a word order constraint. Because the animacy constraint plays no role in production, asymmetry is modeled. The model predicts that sentences in which the subject outranks the object in animacy (S>O) are easier to interpret than sentences in which the object outranks the subject in animacy (S<O); sentences in which the subject and object are equal in animacy (S=O) likely fall somewhere in the middle due to the fact that children are in the process of gradually reranking their constraints. Because the animacy constraint does not help distinguish between forms to use in production, SO word order is predicted to be adhered to by preschoolers. The result is asymmetric development of early word order. Adult speakers of Dutch and English are proposed to have the word order constraint correctly ranked above the animacy constraint, so they are expected to interpret active, reversible sentences out of context as SO in comprehension and to use SO order in production.

The second model within the Optimality Theory framework applies the process of incremental optimization proposed by de Hoop and Lamers (2006). If constraints are evaluated as words are encountered in the linguistic stream, then sentences with inanimate subjects, especially S<O sentences, are predicted to be at a disadvantage during processing.

I discuss in Chapter 3 how the comprehension and production of word order can most effectively be tested in preschoolers. Elicited production was used as a production task, where a child saw an action performed or depicted and had to describe it linguistically. Act-out, picture selection, and preferential looking were used as comprehension tasks. In an act-out task a child had to perform an action with toys that reflects the meaning of a sentence they hear. In a picture selection task the child had to chose the animation (from a pair of animations) that reflects the meaning of a sentence they hear. In preferential looking, participants' eye gaze was measured to see which of two animations they preferred via visual attention upon hearing a sentence. The experimental design was carefully controlled for confounding factors. In order to test the predictions of the two related theoretical models, six experiments were carried out using these methods, each experiment with a comprehension and production component.

### Chapters 4 – 8: Experiments and results

In the experimental chapters, I investigate how Dutch- and English-speaking preschoolers, as well as adult controls, interpret reversible transitive sentences in which the relative animacy of subject and object have been manipulated. For each comprehension task, an accompanying production task determines how the same transitive sentences are formed by the same individuals.

Chapter 4 and Chapter 5 report two aspects of the same experiments using toys. Chapter 4 presents the comprehension data from two act-out tasks with Dutch preschoolers, and Chapter 5 presents production data from two elicitation tasks carried out with the same populations. The first population of Dutch preschoolers was tested with sentences with an animal-vehicle animacy contrast and the second with sentences with a human-vehicle contrast. Chapter 6 and Chapter 7 also report two aspects of the same experiments, this time involving video animations and including measurements from automated eye tracking. Chapter 6 presents comprehension data from a picture selection task with Dutch- and English-speaking adults and preschoolers; the children were additionally tested with a preferential looking task. Chapter 7 presents production data from an elicitation task carried out with the same four populations tested in Chapter 6. All experiments reported in Chapter 6 and Chapter 7 involve only sentences with an animacy-vehicle distinction. Chapter 8 compares performance on the tasks across all of the experiments with regard to scorability and accuracy.

The results show that the Dutch preschoolers tested with act-out and elicitation tasks were able to produce S-O word order more reliably than they were able to comprehend it. Sentences were produced with SO order 95% – 100% of the time on average by the three-and-a-half-year-olds, whereas they had SO interpretations of the same sentences 84% – 91% of the time. The discrepancy appeared to be greater in the younger group of children: the two-and-a-half-year-olds produced SO order 81% – 85% of the time, but exhibited SO interpretations only about 59% – 64% of the time. The accuracy analysis presented in Chapter 8, which accounted for items that received a scorable response by a child on both comprehension and production, confirmed that task type (production vs. comprehension) was indeed a significant predictor of whether SO order would be reflected in a child's response. Namely, a child was more likely to produce word order as SO than he or she was to interpret word order as SO in the act-out paradigm.

Similar results were found with Dutch- and English-speaking preschoolers tested with picture selection and elicitation tasks. The average performance on the production task versus the picture selection task appeared to differ for both the older (92% vs. 70%) and younger (81% vs. 54%) Dutch children. Likewise, the average performance on the production task versus the picture selection task appeared to differ for the older (94% vs. 80%) and the younger (79% vs. 60%) English-speaking children. The accuracy analysis in Chapter 8 confirmed that task (production vs. comprehension) was again a significant predictor in both languages of whether SO order would be reflected in a child's response. A child was more likely to produce SO word order than he or she was to interpret word order as SO in the picture selection paradigm.

The results of the preferential looking task with animations reflected a general comprehension of word order by both age groups of preschoolers in both languages, since as the time progressed over the trial, their looks to the target animation increased. However, the proportion of looking to target never reached above 70% in any condition in either age group in either language. This maximum proportion of looks to SO animations was lower than the average performance on the accompanying production tasks with animations reported in the preceding paragraph.

The results further show that Dutch- and English-speaking adult controls performed similarly to each other on the picture selection task. Adults of both languages interpreted word order as SO 97% of the time on average, indicating a general preference to interpret initial NPs as subjects. However, they were more likely to do so when the subject was animate than when it was inanimate. Likewise, an effect of subject animacy was found in their reaction times, with participants of both languages faster to answer when the subject was animate. With regard to gaze patterns, adults of both languages showed increased looking to the target within three seconds of hearing the sentence subject. For Dutch adults only, this preference for the target was greater in sentences containing an animate subject compared to sentences with an inanimate subject. Thus, while word order was greatly adhered to, there were distinct effects of subject animacy in both the offline answers given by Dutch- and English-speaking adults, as well as in their processing behavior.

## Conclusions

In the final chapters, the results of the experiments are discussed in light of the theoretical and methodological issues raised in Chapters 2 and 3. The discussion in Chapter 9 addresses where predictions were met, where predictions were not met, and possible explanations for discrepancies. In general, the results met the predictions of the OT models: we found an asymmetry between comprehension and production, variable comprehension of word order in Dutch preschoolers, and animacy effects on Dutch- and English-speaking adult controls. Systematic effects of animacy suggest that an animacy constraint is present in both child and adult grammar. Even though variable comprehension was not found in all tasks carried out with preschoolers (with particularly English-speaking children and younger Dutch-speaking children), the asymmetry in favor of production remained. And while animacy had an unexpected influence on word order production of English-speaking children, it did not influence their comprehension. This is seen as further evidence that the two directions of use of grammar do not operate identically.

It is concluded in Chapter 10, based on the results of the four experiments carried out with child populations, that there is an asymmetry between comprehension and production. In two experiments that used toys as stimuli with Dutch preschoolers, and two experiments that used cartoon animations as stimuli with Dutch and English preschoolers, children were more likely to produce word order as SO than they were to interpret word order as SO. We can be confident about the comprehension-production asymmetry found in this investigation because it occurred across different tasks that were adequately controlled for confounding factors, such as verb or likelihood biases. Children's comprehension of word order clearly suffers in the absence of other linguistic and contextual cues that normally aid them in determining agent and patient in natural situations. S-O animacy is rarely the only source of information about agent-patient relations, but may be used by preschoolers when everything else besides word order is stripped away. In contrast, children have little difficulty producing the same sentences that they have difficulty interpreting.

It is additionally concluded that animacy plays a systematic role in sentence interpretation due to an animacy constraint present in the grammar. This is based on the variable comprehension of word order found in Dutch preschoolers as well as animacy affects on Dutch- and English-speaking adult controls. However, variable comprehension was not found in all tasks carried out with preschoolers, with particularly unexpected results from English-speaking children and from children at age 2;6 in general. Furthermore, animacy played a role during the production task, indicating that animacy may be relevant in production, albeit in a different manner than in comprehension.

# Nederlandse Sammenvatting

## Hoofdstuk 1: Het verschijnsel

Wanneer je een kind tussen de 2½ en 3½ jaar oud vraagt om het zinnetje “De auto duwt de koe” uit te beelden met speelgoed, dan zal het in de helft van de gevallen de koe de auto laten duwen (Chan, Lieven & Tomasello, 2009; Chapman & Miller, 1975). Zelfs als je de invloed buiten beschouwing laat van de selectierestricties van het werkwoord en de waarschijnlijkheid van de gebeurtenis, negeren Engelssprekende kinderen regelmatig de woordvolgorde van de zin. Ze vinden het belangrijker dat de door de zin uitgedrukte handeling wordt verricht door een levend iets in plaats van een niet-levend iets.

Dit verschijnsel, dat kleuters geen gebruik maken van woordvolgorde bij hun begrip van omkeerbare transitieve zinnen, is bijzonder interessant vanwege hun ogenschijnlijk volwassen productie van woordvolgorde (Angiolillo & Goldin-Meadow, 1982; Chapman & Miller, 1975; McClellan, Yewchuk & Holdgrafer, 1986). Dit is opvallend, aangezien begrip meestal voorloopt op productie. De mogelijkheid dat kinderen woordvolgorde in een omkeerbare zin zoals hierboven anders gebruiken in begrip dan in productie is de aanleiding voor de twee centrale thema’s in dit proefschrift: asymmetrie en levendheid. Zijn kinderen beter in het produceren van subject-object (SO) woordvolgorde dan in het begrijpen ervan? En als dat zo is, wordt hun begrip systematisch beïnvloed door de relatieve levendheid van het subject en het object?

## Hoofdstuk 2 & 3: Theoretisch kader, methodologie en voorspellingen

Om deze vragen te kunnen beantwoorden, stel ik in hoofdstuk 2 eerst vast welk theoretisch kader het beste in staat is om asymmetrieën tussen productie en begrip bij jonge kinderen te verklaren. Veel theorieën zijn niet in staat om fouten in begrip in combinatie met volwassen productie te verklaren. Optimality Theory kan dat wel, door begrip en productie als twee “richtingen” van dezelfde grammatica te beschouwen. Het hiërarchische systeem van schendbare constraints in Optimality Theory, waarvan sommige symmetrie bevorderen terwijl andere juist asymmetrie bevorderen, kan deze asymmetrie in de vroege ontwikkeling van woordvolgorde verklaren.

Om specifieke voorspellingen te kunnen doen over het gebruik van woordvolgorde in het Nederlands en het Engels door peuters en volwassenen, maak ik gebruik van twee theoretische modellen binnen Optimality Theory. Het eerste model, van Hendriks, De Hoop en Lamers (2005), doet voorspellingen over de uiteindelijke interpretatie van de zin. Hun model maakt gebruik van een levendheidsconstraint die ten onrechte boven een woordvolgordeconstraint gerangschikt is. Omdat deze levendheidsconstraint geen rol speelt in productie, kan een asymmetrie ontstaan tussen productie en begrip. Dit model voorspelt dat zinnen met een subject dat hoger is in levendheid dan het object (S>O) makkelijker te interpreteren zijn dan zinnen met een object dat hoger is in levendheid dan het subject (S<O). Zinnen waarin het subject en het object allebei levend of allebei niet-levend zijn (S=O) vallen hier waarschijnlijk tussenin, omdat kinderen nog bezig zijn met het geleidelijk herrangschikken van hun constraints. Omdat de levendheidsconstraint niet helpt bij de keuze tussen verschillende woordvolgordes, voorspelt dit model een SO-woordvolgorde voor peuters in productie. Ten gevolge hiervan voorspelt het model dat de vroege ontwikkeling van woordvolgorde bij kinderen asymmetrisch verloopt en dat productie voorloopt op begrip. Aangenomen wordt dat volwassen sprekers van het Nederlands en het Engels hun woordvolgordeconstraint boven hun levendheidsconstraint hebben gerangschikt. Daardoor

kennen zij aan actieve, omkeerbare zinnen zonder enige context een SO-interpretatie toe, en gebruiken zij SO-woordvolgorde in hun taalproductie.

Het tweede model binnen het theoretische kader van OT maakt gebruik van het proces van incrementele optimalisatie, zoals voorgesteld door De Hoop en Lamers (2006). Als taalgebruikers de constraints van hun grammatica toepassen bij de woord-voor-woord interpretatie van de zin , dan wordt op basis van dit model verwacht dat zinnen met een niet-levend subject—met name S<O zinnen—meer verwerkingsmoeilijkheden opleveren dan zinnen met een levend subject.

In hoofdstuk 3 bespreek ik hoe begrip en productie van woordvolgorde het meest effectief kunnen worden getest bij peuters. Om productie te testen gebruik ik een productietaak waarbij het kind een actie ziet die uitgebeeld wordt met speelgoed. Het kind moet deze actie vervolgens in woorden beschrijven. Als begripstaak gebruik ik verschillende soorten taken en methoden: een uitbeeldingstaak, een plaatjes-selectie-taak en het meten van oogbewegingen. Bij de uitbeeldingstaak moet het kind een actie uitvoeren met speelgoed op basis van een gehoorde zin. Bij de plaatjes-selectie-taak moet het kind uit twee tekenfilmpjes het tekenfilmpje kiezen dat het beste past bij de gehoorde zin. Bij de oogbewegingstaak wordt het kijkgedrag van het kind gemeten, om te zien op welk moment het kind aan welk van de twee tekenfilmpjes de voorkeur geeft tijdens het luisteren naar de zin. De onderzoeksopzet is zorgvuldig gecontroleerd en gebalanceerd voor wat betreft mogelijke taakinvloeden. Om de voorspellingen van de twee modellen te testen worden er in totaal zes experimenten uitgevoerd. Al deze experimenten bestaan uit een begripsgedeelte en een productiegedeelte.

## Hoofdstuk 4 – 8: Experimenten en resultaten

In de hoofdstukken 4 t/m 7 onderzoek ik hoe Nederlandssprekende en Engelssprekende peuters en volwassenen omkeerbare transitieve zinnen interpreteren. In deze zinnen heb ik de relatieve levendheid van het subject en het object gemanipuleerd. Elke begripstaak is gekoppeld aan een vergelijkbare productietaak, zodat we kunnen testen hoe dezelfde omkeerbare transitieve zinnen door dezelfde proefpersonen worden geproduceerd.

In hoofdstuk 4 en 5 worden twee aspecten beschreven van dezelfde experimenten met speelgoed. In hoofdstuk 4 worden de begripsresultaten gepresenteerd van twee uitbeeldingstaken met Nederlandse peuters. In hoofdstuk 5 worden de productieresultaten gepresenteerd van twee productietaken bij dezelfde groep kinderen. De eerste groep Nederlandse peuters is getest met zinnen waarin het contrast in levendheid werd uitgebeeld door dier-voertuig-combinaties. Bij de tweede groep Nederlandse peuters werd het contrast in levendheid uitgebeeld door mens-voertuig-combinaties.

In hoofdstuk 6 en 7 worden eveneens twee aspecten van dezelfde experimenten gepresenteerd, in dit geval de experimenten met tekenfilmpjes. In hoofdstuk 6 worden de begripsresultaten besproken van een plaatjes-selectie-taak met Nederlandssprekende en Engelssprekende volwassenen en peuters. Deze kinderen werden tevens getest met behulp van een oogbewegingstaak. In hoofdstuk 7 worden de productieresultaten besproken van dezelfde groepen proefpersonen als in hoofdstuk 6. Alle experimenten in hoofdstuk 6 en 7 maken gebruik van zinnen met een dier-voertuig-contrast.

In hoofdstuk 8 worden de resultaten van alle experimenten uit voorgaande hoofdstukken vergeleken. Hierbij wordt in het bijzonder gekeken naar de bruikbaarheid en correctheid van de antwoorden. De resultaten laten zien dat de Nederlandse peuters—getest met een uitbeeldingstaak en een productietaak—beter in staat zijn om SO-woordvolgorde te produceren dan te begrijpen. In 95% tot 100% van de gevallen produceerden de 3½-jarigen zinnen met SO-volgorde, terwijl ze dergelijke SO-zinnen in slechts 84% tot 91% van de gevallen correct interpreteerden. Dit verschil was nog groter in de jongste groep kinderen: de 2½-jarigen produceerden in 81% tot 85% van de gevallen zinnen met SO-volgorde, terwijl ze deze zinnen in slechts 59% tot 64% van de gevallen correct interpreteerden. De analyse van correcte antwoorden in hoofdstuk 8 (waarin rekening wordt gehouden met bruikbare antwoorden op zowel de productietaak als de begripstaak) bevestigt dat het type taak—productie of begrip—een betrouwbare voorspeller is van het gebruik van SO-woordvolgorde. In de uitbeeldingstaak was het waarschijnlijker dat een kind SO-woordvolgorde produceerde dan dat het kind een SO-interpretatie uitbeeldde voor een gehoorde zin.

Vergelijkbare resultaten werden gevonden in experimenten waarin Nederlandse en Engelse peuters getest werden met een plaatjes-selectie-taak en een productietaak. Gemiddeld deden de kinderen het beter in de productietaak dan in de plaatjes-selectie-taak. Dit gold voor zowel de oudste groep Nederlandse kinderen (92% tegenover 70%) als de jongste groep Nederlandse kinderen (81% tegenover 54%). Hetzelfde patroon werd gevonden voor de Engelse kinderen. De oudste groep Engelse kinderen was beter in productie dan in begrip (94% tegenover 80%), evenals de jongste groep Engelse kinderen (79% tegenover 60%). Opnieuw bleek uit de analyse van correcte antwoorden in hoofdstuk 8 dat het type taak—productie of begrip—in beide talen een goede voorspeller is voor het gebruik van SO-woordvolgorde. In de plaatjes-selectie-taak was het namelijk waarschijnlijker dat een kind SO-woordvolgorde produceerde dan dat het kind een SO-interpretatie koos.

De resultaten van de oogbewegingstaak met tekenfilmpjes lieten zien dat kinderen in beide leeftijdsgroepen in beide talen woordvolgorde over het algemeen goed begrepen: in de loop van een trial keken de kinderen steeds meer naar het correcte filmpje. Desalniettemin kwam de proportie van blikken naar het correcte filmpje nooit boven de 70% uit, in geen van de condities, leeftijdsgroepen of talen. Deze maximale proportie van blikken naar de SO-tekenfilmpjes was lager dan de proportie van correct geproduceerde SO-zinnen in de bijbehorende productietaken.

De resultaten laten verder zien dat Nederlandssprekende en Engelssprekende volwassenen zeer vergelijkbaar presteerden op de plaatjes-selectie-taak. Volwassenen in beide talen interpreteerden woordvolgorde als SO-woordvolgorde in gemiddeld 97% van de gevallen. Dit wijst op een algemene voorkeur om NPs in eerste zinspositie als subjecten te interpreteren. Deze voorkeur was sterker wanneer het subject levend was dan wanneer het subject niet-levend was. Dit bleek ook uit de reactietijden van de proefpersonen: proefpersonen antwoordden sneller wanneer het subject levend was dan wanneer het subject niet-levend was. Met betrekking tot de oogbewegingen valt op dat beide groepen volwassenen binnen drie seconden nadat ze het subject hoorden, naar het correcte filmpje keken. Alleen de Nederlandstalige volwassenen hadden een lichte voorkeur voor levende subjecten in vergelijking met niet-levende subjecten. Samenvattend kunnen we stellen dat, hoewel Nederlandssprekende en Engelssprekende volwassenen zich over het algemeen hielden aan de woordvolgorde, ze duidelijke effecten lieten zien van levendheid in zowel hun uiteindelijke interpretatie van de zin als tijdens de verwerking van de zin.

## Conclusies

In de laatste hoofdstukken worden de experimentele resultaten besproken in relatie tot de theorie en de methoden zoals uiteengezet in hoofdstuk 2 en 3. In de discussie in hoofdstuk 9 bespreek ik welke voorspellingen worden bevestigd door de resultaten, en welke voorspellingen niet worden bevestigd door de resultaten. Ook worden mogelijke verklaringen gegeven voor de gevallen waarin de voorspellingen niet uitkwamen. Over het algemeen kwamen de resultaten overeen met de voorspellingen van de OT-modellen: we vonden een asymmetrie tussen begrip en productie, een niet-volwassen begrip van woordvolgorde door Nederlandse peuters en effecten van levendheid bij Nederlandssprekende en Engelssprekende volwassenen. Systematische effecten van levendheid duiden op het bestaan van een levendheidsconstraint in de grammatica van kinderen en volwassen. Hoewel we niet in alle taken met peuters (met name niet in de taken met de Engelse kinderen en de jongste Nederlandse kinderen) variatie in het begrip van woordvolgorde vonden, was er wel sprake van een asymmetrie, met productie voorlopend op begrip. Levendheid had een onverwacht effect op de productie van woordvolgorde door de Engelse kinderen, maar had geen invloed op hun begrip. Dit kan worden gezien als een aanwijzing dat het effect van de grammatica afhangt van de richting van toepassing: productie of begrip.

In hoofdstuk 10 wordt op basis van de vier experimenten met kinderen geconcludeerd dat er een asymmetrie is tussen begrip en productie. In de twee experimenten waarin speelgoed werd gebruikt om Nederlandse peuters te testen, en de twee experimenten waarin tekenfilmpjes werden gebruikt om Nederlandse en Engelse peuters te testen, lieten de kinderen vaker correcte productie van SO-woordvolgorde zien dan correcte interpretatie van SO-woordvolgorde. Dit effect is robuust en betrouwbaar, aangezien het naar voren kwam in verschillende experimenten waarin voor de eventuele invloed van werkwoorden en waarschijnlijkheid was gecontroleerd. Kinderen hebben duidelijk moeite met het begrip van woordvolgorde wanneer talige en contextuele aanwijzingen ontbreken die hen in normale situaties helpen bij het vaststellen van wat de agent is en wat de patient is. De levendheid van subject en object is bijna nooit de enige aanwijzing voor de agent-patient-relatie in een zin, maar wordt gebruikt door peuters als er geen andere aanwijzingen beschikbaar zijn. Daarentegen hebben kinderen weinig moeite om diezelfde zinnen te produceren.

Verder kan worden geconcludeerd dat levendheid een systematische rol speelt bij het interpreteren van zinnen, wellicht doordat er een levendheidsconstraint aanwezig is in de grammatica. We baseren deze conclusie op de fouten die Nederlandse peuters maken in hun begrip van woordvolgorde, en op de effecten van levendheid bij Nederlandse en Engelse volwassenen. We vonden echter niet in alle taken met peuters een effect van levendheid, in het bijzonder niet bij de Engelssprekende kinderen en bij kinderen van 2½ jaar in het algemeen. Levendheid speelde eveneens een rol in de productietaken. Dit duidt erop dat levendheid weliswaar van belang is in productie, maar op een andere manier dan in begrip.

# Groningen Dissertations in Linguistics (GRODIL)

1. Henriëtte de Swart (1991). *Adverbs of Quantification: A Generalized Quantifier Approach.*
2. Eric Hoekstra (1991). *Licensing Conditions on Phrase Structure.*
3. Dicky Gilbers (1992). *Phonological Networks. A Theory of Segment Representation.*
4. Helen de Hoop (1992). *Case Configuration and Noun Phrase Interpretation.*
5. Gosse Bouma (1993). *Nonmonotonicity and Categorial Unification Grammar.*
6. Peter I. Blok (1993). *The Interpretation of Focus.*
7. Roelien Bastiaanse (1993). *Studies in Aphasia.*
8. Bert Bos (1993). *Rapid User Interface Development with the Script Language Gist.*
9. Wim Kosmeijer (1993). *Barriers and Licensing.*
10. Jan-Wouter Zwart (1993). *Dutch Syntax: A Minimalist Approach.*
11. Mark Kas (1993). *Essays on Boolean Functions and Negative Polarity.*
12. Ton van der Wouden (1994). *Negative Contexts.*
13. Joop Houtman (1994). *Coordination and Constituency: A Study in Categorial Grammar.*
14. Petra Hendriks (1995). *Comparatives and Categorial Grammar.*
15. Maarten de Wind (1995). *Inversion in French.*
16. Jelly Julia de Jong (1996). *The Case of Bound Pronouns in Peripheral Romance.*
17. Sjoukje van der Wal (1996). *Negative Polarity Items and Negation: Tandem Acquisition.*
18. Anastasia Giannakidou (1997). *The Landscape of Polarity Items.*
19. Karen Lattewitz (1997). *Adjacency in Dutch and German.*
20. Edith Kaan (1997). *Processing Subject-Object Ambiguities in Dutch.*
21. Henny Klein (1997). *Adverbs of Degree in Dutch.*
22. Leonie Bosveld-de Smet (1998). *On Mass and Plural Quantification: The case of French ‘des’/‘du’-NPs.*
23. Rita Landeweerd (1998). *Discourse semantics of perspective and temporal structure.*
24. Mettina Veenstra (1998). *Formalizing the Minimalist Program.*
25. Roel Jonkers (1998). *Comprehension and Production of Verbs in aphasic Speakers.*
26. Erik F. Tjong Kim Sang (1998). *Machine Learning of Phonotactics.*
27. Paulien Rijkhoek (1998). *On Degree Phrases and Result Clauses.*
28. Jan de Jong (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure.*
29. H. Wee (1999). *Definite Focus.*
30. Eun-Hee Lee (2000). *Dynamic and Stative Information in Temporal Reasoning: Korean tense and aspect in discourse.*
31. Ivilin P. Stoianov (2001). *Connectionist Lexical Processing.*
32. Klarien van der Linde (2001). *Sonority substitutions.*
33. Monique Lamers (2001). *Sentence processing: using syntactic, semantic, and thematic information.*
34. Shalom Zuckerman (2001). *The Acquisition of "Optional" Movement.*
35. Rob Koeling (2001). *Dialogue-Based Disambiguation: Using Dialogue Status to Improve Speech Understanding.*
36. Esther Ruigendijk (2002). *Case assignment in Agrammatism: a cross-linguistic study.*

37. Tony Mullen (2002). *An Investigation into Compositional Features and Feature Merging for Maximum Entropy-Based Parse Selection.*
38. Nanette Bienfait (2002). *Grammatica-onderwijs aan allochtone jongeren.*
39. Dirk-Bart den Ouden (2002). *Phonology in Aphasia: Syllables and segments in level-specific deficits.*
40. Rienk Withaar (2002). *The Role of the Phonological Loop in Sentence Comprehension.*
41. Kim Sauter (2002). *Transfer and Access to Universal Grammar in Adult Second Language Acquisition.*
42. Laura Sabourin (2003). *Grammatical Gender and Second Language Processing: An ERP Study.*
43. Hein van Schie (2003). *Visual Semantics.*
44. Lilia Schürcks-Grozeva (2003). *Binding and Bulgarian.*
45. Stasinos Konstantopoulos (2003). *Using ILP to Learn Local Linguistic Structures.*
46. Wilbert Heeringa (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance.*
47. Wouter Jansen (2004). *Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology.*
48. Judith Rispens (2004). *Syntactic and phonological processing in developmental dyslexia.*
49. Danielle Bougaïré (2004). *L'approche communicative des campagnes de sensibilisation en santé publique au Burkina Faso: Les cas de la planification familiale, du sida et de l'excision.*
50. Tanja Gaustad (2004). *Linguistic Knowledge and Word Sense Disambiguation.*
51. Susanne Schoof (2004). *An HPSG Account of Nonfinite Verbal Complements in Latin.*
52. M. Begoña Villada Moirón (2005). *Data-driven identification of fixed expressions and their modifiability.*
53. Robbert Prins (2005). *Finite-State Pre-Processing for Natural Language Analysis.*
54. Leonoor van der Beek (2005) *Topics in Corpus-Based Dutch Syntax.*
55. Keiko Yoshioka (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse.*
56. Sible Andringa (2005). *Form-focused instruction and the development of second language proficiency.*
57. Joanneke Prenger (2005). *Taal telt! Een onderzoek naar de rol van taalvaardigheid en tekstbegrip in het realistisch wiskundeonderwijs.*
58. Neslihan Kansu-Yetkiner (2006). *Blood, Shame and Fear: Self-Presentation Strategies of Turkish Women's Talk about their Health and Sexuality.*
59. Mónika Z. Zempléni (2006). *Functional imaging of the hemispheric contribution to language processing.*
60. Maartje Schreuder (2006). *Prosodic Processes in Language and Music.*
61. Hidetoshi Shiraishi (2006). *Topics in Nivkh Phonology.*
62. Tamás Biró (2006). *Finding the Right Words: Implementing Optimality Theory with Simulated Annealing.*
63. Dieuwke de Goede (2006). *Verbs in Spoken Sentence Processing: Unraveling the Activation Pattern of the Matrix Verb.*
64. Eleonora Rossi (2007). *Clitic production in Italian agrammatism.*
65. Holger Hopp (2007). *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing.*

66. Gerlof Bouma (2008). *Starting a Sentence in Dutch: A corpus study of subject- and object-fronting.*
67. Julia Klitsch (2008). *Open your eyes and listen carefully. Auditory and audiovisual speech perception and the McGurk effect in Dutch speakers with and without aphasia.*
68. Janneke ter Beek (2008). *Restructuring and Infinitival Complements in Dutch.*
69. Jori Mur (2008). *Offline Answer Extraction for Question Answering.*
70. Lonneke van der Plas (2008). *Automatic Lexico-Semantic Acquisition for Question Answering.*
71. Arjen Versloot (2008). *Mechanisms of Language Change: Vowel reduction in 15th century West Frisian.*
72. Ismail Fahmi (2009). *Automatic term and Relation Extraction for Medical Question Answering System.*
73. Tuba Yarbay Duman (2009). *Turkish Agrammatic Aphasia: Word Order, Time Reference and Case.*
74. Maria Trofimova (2009). *Case Assignment by Prepositions in Russian Aphasia.*
75. Rasmus Steinkrauss (2009). *Frequency and Function in WH Question Acquisition. A UsageBased Case Study of German L1 Acquisition.*
76. Marjolein Deunk (2009). *Discourse Practices in Preschool. Young Children's Participation in Everyday Classroom Activities.*
77. Sake Jager (2009). *Towards ICT-Integrated Language Learning: Developing anImplementation Framework in terms of Pedagogy, Technology and Environment.*
78. Francisco Dellatorre Borges (2010). *Parse Selection with Support Vector Machines.*
79. Geoffrey Andogah (2010). *Geographically Constrained Information Retrieval*.
80. Jacqueline van Kruiningen (2010). *Onderwijsontwerp als conversatie. Probleemoplossingin interprofessioneel overleg.*
81. Robert G. Shackleton (2010). *Quantitative Assessment of English-American SpeechRelationships.*
82. Tim Van de Cruys (2010). *Mining for Meaning: The Extraction of Lexico-semantic Knowledge from Text.*
83. Therese Leinonen (2010). *An Acoustic Analysis of Vowel Pronunciationin Swedish Dialects.*
84. Erik-Jan Smits (2010). *Acquiring Quantification. How Children Use Semantics andPragmatics to Constrain Meaning.*
85. Tal Caspi (2010). *A Dynamic Perspective on Second Language Development.*
86. *Teodora Mehotcheva (2010). After the fiesta is over. Foreign language attrition of Spanish in Dutch and German Erasmus Student.*
87. Xiaoyan Xu (2010). *English language attrition and retention in Chinese and Dutchuniversity students.*
88. Jelena Prokic (2010). *Families and Resemblances.*
89. Radek Šimík (2011). *Modal existential wh-constructions.*
90. Katrien Colman (2011). *Behavioral and Neuroimaging studies on language processing in Dutch speakers with Parkinson's disease.*
91. Siti Mina Tamah (2011). *Student interaction in the implementation of the jigsaw technique in language teaching.*
92. Aletta Kwant (2011). *Geraakt door prentenboeken. Effecten van het gebruik van prentenboeken op de sociaal-emotionele ontwikkeling van kleuters.*

93. Marlies Kluck (2011). *Sentence amalgamation.*
94. Anja Schüppert (2011). *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish.*
95. Peter Nabende (2011). *Applying Dynamic Bayesian Networks in Transliteration Detection and Generation.*
96. Barbara Plank (2011). *Domain Adaptation for Parsing.*
97. Çağrı Çöltekin (2011). *Catching Words in a Stream of Speech: Computational simulations of segmenting transcribed child-directed speech.*
98. Dörte Hessler (2011). *Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners: The Whole is More than the Sum of its Parts*
99. Herman Heringa (2012). *Appositional constructions.*
100. Diana Dimitrova (2012). *Neural Correlates of Prosody and Information Structure.*
101. Harwintha Anjarningsih (2012). *Time Reference in Standard Indonesian Agrammatic Aphasia.*
102. Myrte Gosen (2012). *Tracing learning in interaction. An analysis of shared reading of picture books at kindergarten.*
103. Martijn Wieling (2012). *A Quantitative Approach to Social and Geographical Dialect Variation.*
104. Gisi Cannizzaro (2012). *Early word order and animacy.*

GRODIL
Center for Language and Cognition Groningen (CLCG)
P.O. Box 716
9700 AS Groningen
The Netherlands

# Curriculum Vitae

Gisi Cannizzaro studied German language and culture at Loyola University in New Orleans, graduating magna cum laude with a bachelor degree in 2004. She continued her studies in 2006 as a student of the European Masters in Clinical Linguistics Erasmus study program at the University of Groningen, the University of Joensuu, and the University of Potsdam. Under the supervision of Prof. Barbara Höhle, Gisi completed her masters thesis addressing verb-object word order produced by German- and English-speaking two-year-olds. In 2008 she began her PhD project on word order in early child language at the Center for Language and Cognition at the University of Groningen. Her PhD research is part of Prof. Petra Hendriks' NWO Vici project "Asymmetries in Grammar." In order to expand on her knowledge about eye tracking with children, she visited Prof. John Trueswell's language lab at the University of Pennsylvania's Institute for Research in Cognitive Science in 2010 and 2011. Gisi helped organize numerous international conferences and workshops, regularly gave guest lectures for courses on language acquisition and psycholinguistics, and served four years as eye lab coordinator for the Asymmetries project.